

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
11 January 2001 (11.01.2001)

PCT

(10) International Publication Number
WO 01/02856 A1

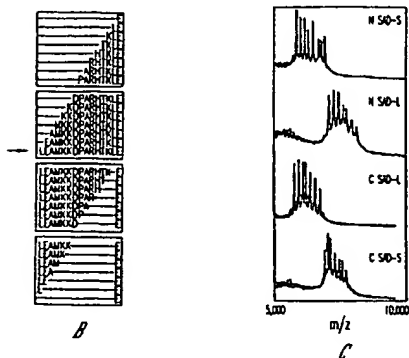
- (51) International Patent Classification⁷: G01N 33/53, C07K 14/00, C12N 15/11
- (71) Applicant and
(72) Inventor: WINSTON, Rachel, L. [US/US]; 1451 Shotwell Street, San Francisco, CA 94110 (US).
- (21) International Application Number: PCT/US00/18335
- (74) Agent: MCCrackin, Ann, M.; Schwegman, Lundberg, Woessner & Kluth, P.O. Box 2938, Minneapolis, MN 55402 (US).
- (22) International Filing Date: 3 July 2000 (03.07.2000)
- (25) Filing Language: English
- (81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (26) Publication Language: English
- (30) Priority Data: 60/142,259 2 July 1999 (02.07.1999) US
- (71) Applicant (*for all designated States except US*): THE SCRIPPS RESEARCH INSTITUTE [US/US]; 10550 North Torrey Pines Road, LaJolla, CA 92037 (US).
- (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European

[Continued on next page]

(54) Title: METHOD FOR DETERMINATION OF PEPTIDE-BINDING AGENT INTERACTION

39 50 60 70 80 90 107
ELRKTNKPKMKRPRRARRHCLNLSKSL LEAMAKDPAHITKLE KADLEMTYKRLQSVORQOL
HSE HSE HSE HSE

A



(57) Abstract: The present invention relates to methods for analyzing, altering, and controlling the structural basis for protein binding to target molecules using modified polypeptide libraries. More particularly, the present invention is directed to addition, deletion, and substitution modified polypeptide libraries and to the use of modified polypeptide libraries for probing the interactions between proteins and other molecules.

WO 01/02856 A1



patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

Published:

— *With international search report.*

METHOD FOR DETERMINATION OF PEPTIDE – BINDING AGENT INTERACTION

Government Funding

5 The invention described herein was made with government support under Grant Numbers GM-26453, GM-47530, and GM-57148 awarded by the National Institute of Health. The United States Government has certain rights in the invention.

Cross-Reference To Related Applications

10 The present application is based upon, incorporates the disclosure and inventions described in, and claims priority from U.S. provisional patent application serial no. 60/142,259, filed July 2, 1999.

Background of the Invention

Two general research approaches have emerged to accurately probe
15 interactions between proteins and other molecules. In the context of enzyme-substrate research, an early approach focused on covalently modifying the structures of substrates, while holding as constant the structures and properties of enzymes. Alan Fersht, Enzyme Structure and Mechanism, New York: Freeman (2nd ed., 1985). For example, modifications in substrate electronegativity and
20 chain length were used to elucidate the reaction kinetics of enzymes such as the serine proteases chymotrypsin, elastase, and pepsin. In a related approach, transition state analogues were used to study the reaction kinetics of a variety of enzymes, such as lysozyme, proline racemase, and cytidine deaminase.

In contrast to the first approach, a later approach focused on covalently
25 modifying the structures of proteins, while holding as constant the structures and properties of acceptor molecules such as substrates. Amino acid residues in an enzyme may be changed in a systematic manner by using site-directed mutagenesis. Mutant enzymes can be prepared that, for instance, lack sidechains that are necessary to bind substrates. As a result, the effect of the modification
30 on binding energy and catalysis can be measured. The first study adopting this approach employed the tyrosyl-tRNA synthetases, a class of enzymes that catalyze the aminoacylation of tRNA. The technique of site-directed

mutagenesis has made it possible to probe the effects of individual sidechains on protein function in a number of instances. This approach to studying protein structure and function can be tedious, however, because modified proteins must be prepared and tested one at a time.

- 5 Covalently modified proteins have also been prepared by classical chemical synthesis techniques. For example, N-methylation and the use of ester bonds can probe backbone interactions (Arad et al. Biopolymers 1990, 29, 1633-1649; Bramson et al. J. Biol. Chem. 1985, 260, 15452-15457; Caporale et al. In: Peptides: Structure and Function, Proceedings of the Tenth American Peptide
10 Symposium; Marshall, G.F. Ed. Escom: Leiden: The Netherlands, 1988, pp. 449-451), while sidechain contributions can be probed using D-amino acid or Alanine/Glycine substitutions (Konishi et al. In: Peptides: Structure and Function, Proceedings of the Tenth American Peptide Symposium, Marshall, G.F. Ed. Escom, Leiden: The Netherlands, 1988, pp. 479-481; Tam et al. In
15 Peptides: Proceedings of the Eleventh American Peptide Symposium; Rivier, J. E.; Marshall, G. R. Ed.; Escom: Leiden, The Netherlands, 1990, pp. 75-77). As traditionally practiced, a separate analogue must be prepared and assayed for each position in the peptide sequence that is to be studied.

- An alternative method of studying peptides is through combinatorial
20 chemistry. This approach has had a noteworthy impact on the study of the molecular basis of peptide activity and has contributed to the search for new biologically active peptides (Thompson et al. Chem. Rev. 1996, 96, 555-600; Gordon et al. J. Med. Chem. 1994, 37, 1385-1401; Scott et al. Curr. Op. Biotech 1994, 5, 40-48). 'Multiple Peptide Synthesis' has extended the traditional
25 approach by allowing multiple peptides to be synthesized simultaneously (Geysen et al. J. Proc. Natl. Acad. Sci. USA 1984, 81, 3998-4001; Houghten et al. Proc Natl. Acad. Sci. USA 1985, 82, 5131-5134). The individual peptide products are spatially separated and can be analyzed either attached to a solid support or in solution. Established 'split synthesis' (Furka et al. Int. LT. Pept.
30 Prot. Res. 1991, 37, 487-494; Lam et al. Nature 1991, 354, 82-84) procedures allow for the rapid generation of large numbers of peptide sequences through the repetition of a simple divide, couple and recombine process.

The compositional diversity made possible by the combinatorial chemistry approach is advantageous for the discovery of new 'lead' compounds because, in principle, all possible structural variants can be explored for the desired activity and only the few active polypeptides of interest need to be
5 individually identified (Furka et al. *Int. J. Pept. Prot. Res.* 1991, 37, 487-494; Lam et al. *Nature* 1991, 354, 82-84). Such libraries may be too complex to fully characterize and may have limited utility where information about a complete set of functional and non-functional components is desired over many positions in a peptide sequence.

10 A more systematic investigation of the molecular basis of peptide function requires a different type of molecular diversity. Instead of a peptide mixture of high compositional diversity, it would be useful to construct an array of peptides, which differ from each other in a precise and defined manner. In principle, one way to access this population would be as a minor fraction of a
15 large, fully combinatorial library. For example, such an array of analogues could consist of all peptides that differ from a target sequence by a single amino acid substitution at each position in a peptide sequence (cf. 'Ala scans'). By removing this defined subset of analogues from the context of a complex, fully combinatorial mixture of peptides, handling and analysis would be greatly
20 simplified and a more useful profile of the effects of substituting the amino acid throughout the peptide chain would be obtained. Current split resin methods do not allow for this type of control over the composition of a peptide library. (Furka et al. *Int. J. Pept. Prot. Res.* 1991, 37, 487-494; Lam et al. *Nature* 1991, 354, 82-84).

25 Typically, to investigate the molecular basis of protein function, systematic modifications are made to the protein structure and the effects of those modifications on the properties of the protein are evaluated. Site-directed mutagenesis (Smith et. . al. *Angew. Chem. Int. Ed. Engl.* 1994, 33, 1214-1220) has been the principle tool used to implement this approach and has given many
30 insights into the contribution of individual sidechains to protein function. In particular, 'alanine scanning' (Wells et al. *Methods in Enzymology* 1991, 202, 390-411) has been used to identify specific amino acid sidechains involved in ligand binding interactions. This technique involves the sequential substitution

of native amino acids by individual alanine residues, which are regarded as functionally and structurally neutral. To extend the repertoire of modifications beyond the twenty genetically encoded amino acids, methods have been developed to substitute non-natural groups into proteins (Noren et al. Science. 5 1989, 244, 182-185). Although a variety of both novel sidechain and backbone modified proteins have been generated, there are apparent limits to the modifications possible using the methods of molecular biology and ribosomal synthesis (Ellman et al. Science 1991, 255, 197-200; Cornish et al. Angew Chem Int. Ed. Engl. 1995, 34, 621-633). Recent advances in the total synthesis of 10 polypeptides have opened the world of proteins to direct application of the tools of organic chemistry (Schnolzer et al. Science 1992, 256, 221-225; Jackson et al. Science 1994, 266, 243-247; Dawson et al. Science 1994, 266, 776-779; Canne et al. J. Am. Chem. Soc. 1995, 117, 2998-3007; Liu et al. J. Am. Chem. Soc. 1995, 118, 307-312; Englebretsen et al. Tet. Lett. 1995, 36, 8871-8874). Using 15 total chemical synthesis, a variety of protein analogues have been synthesized. Of particular note have been proteins containing B-turn mimics (Baca et al. Prot. Sci. 1993, 2, 1085-1091), N-methylated amino acids (Rajaratnam et al. Science 1994, 264, 90-92), modified backbone atoms (Baca et al. J. Am. Chem. Soc. 1995, 117, 1881-1887), and mirror image proteins composed entirely of D- 20 amino acids (Zawadzke et al. J. Am. Chem. Soc. 1992, 114, 4002-4003; Milton et al. Science 1992, 256, 1445-1448; Fitzgerald et al. J. Am. Chem. Soc. 1995, 117, 11075-11080; Schumaacher et al. Science 1996, 271, 1854-1857). In addition, important insights into the mechanism of action of enzymes have been attained through the total chemical synthesis of unique analogues (Baca et al. 25 Proc. Natl. Acad. Sci. U.S.A. 1993, 90, 11638-11642).

Although structure-function relationships in proteins can be studied using individual analogues prepared by either recombinant or chemical techniques, development of a profile of effects across the whole protein molecule is hindered by the time and effort required to generate and analyze multiple protein 30 analogues (Matthews et al. Ann. Rev. Biochem. 1993, 62, 139-160). The use of combinatorial oligonucleotide synthesis in conjunction with protein expression in bacteria (Reidhaar-Olsen et al. Science 1988, 241, 53-57; Gregoret et al. Proc. Natl. Acad. Sci. USA. 1993, 90, 4246-4250) or on phage (Scott et al. Science

1990, 249, 386-390; Lowman, H. B. Bass, S.H.; Simpson, N.; Wells, J. A. Biochemistry 1991. 30 10832-10838) has provided a powerful method for studying large numbers of analogue proteins. These techniques allow pools of expressed proteins to be probed for a desired function. With appropriate
5 screening procedures, a statistical sampling of numerous functional protein variants can be analyzed and identified (Gu et al. Protein Science 1995, 4, 1108-1117). This strategy has proved to be powerful for generating variant proteins with new or optimized functions (Lowman et al. J. Moll. Biol. 1993, 234, 564-578; Rebar et al. Science 1994, 263, 671-673). However, approaches designed
10 to elucidate the molecular basis of protein function have been complicated by the necessarily incomplete characterization of the numerous protein analogues generated. Some of the mutant proteins contained in libraries may ultimately not be expressed because they are fatal to the bacteria used in protein expression. Studies are also hampered by necessary limitation to the naturally encoded
15 amino acids.

Recently, the valuable information that can be gained by systematic modification through chemical synthesis has been combined by researchers with the advantages of combinatorial methods in an approach known as "protein signature analysis." In protein signature analysis, an array of self-encoded
20 protein segments is prepared using the technique of total chemical synthesis. An analogue unit is systematically placed throughout a region of interest in the peptide chain, so that each member of the array contains a single copy of the analogue unit at a unique and defined position. The array of synthetic protein segments containing an analogue unit is then subjected to a selection based on a
25 functional property, such as binding with a substrate or acceptor molecule. This results in a division of the original mixture of peptide segments contained in the peptide array into a positive (functional) pool and a negative (non-functional pool). In the third step of the process, the identities of the synthetic peptide molecules are determined. The position of the analogue unit within each peptide
30 segment is determined using a chemical readout system expressly built into the molecule for that purpose. The resulting patterns form a signature relating to the chemical structure of the molecule to effects on protein function. Muir, et al.

Chem. Biol. 1996 3: 817-825 (1996); Dawson, et al. J. Am. Chem. Soc. 1997 119: 7197-7927; WO 97/11958.

The protein signature analysis technique is useful because it combines the versatility of chemical synthesis for systematically modifying a protein's covalent structure with the practical convenience of combinatorial methods. In one study, the technique was used to probe the chemical basis of binding activity in the SH3 domain. However, the analogue unit that was incorporated into each synthetic peptide contained in the peptide library in the study is the dipeptide Gly-S β Ala, corresponding to $-\text{NHCH}_2\text{COSHCH}_2\text{CH}_2\text{CO}-$. The thioester moiety in Gly-S β Ala is reactive and difficult to use in experimental practice because it readily hydrolyzes. In addition, Gly-S β Ala contains an extra methylene unit (compared to the natural amino acid dipeptide), which may affect the conformation of a synthetic protein and its ability to interact with acceptor molecules. Whether polypeptides containing the Gly-S β Ala linker are good binding site models for other native proteins was left undetermined, largely because the study did not show that the SH3 synthetic peptides prepared by the protein signature analysis technique adopted the correct tertiary structure of the native SH3 domain. Moreover, the study did not individually characterize each synthetic protein. Thus, it has not been determined whether the proteins or protein domains containing synthetic peptide segments as substitutes for native binding sequences are conformationally related to native systems and possess appropriate binding activities.

Therefore, there remains a need identify analogue units that are easy to manipulate. There is an additional need to identify analogue units that mimic natural units. There is a further need to incorporate synthetic peptide segments that contain analogue units or other amino acid additions, substitutions, or deletions into native proteins to study the interactions of proteins with acceptor molecules. There is still a further need for a rapid method for testing the binding activity of native proteins containing synthetic peptide segments to determine functionally important residues of the native protein. There is yet a further need to establish the applicability of this technique to conformationally constrained proteins.

Summary of The Invention

These and other needs are met by the present invention which is directed to a method for determining the interaction between a polypeptide and a binding agent. The method of the invention provides for a systematic analysis of a binding site of a polypeptide such as an enzyme, a receptor, an antibody, a transcription factor and the like. It also provides a method for probing the participation of amino acids in binding. The method enables rapid analysis and is useful for large and small polypeptides, preferably polypeptides with tertiary structures that resemble the tertiary structures of native proteins, hereinafter referred to as "conformationally constrained polypeptides."

The invention includes several aspects involving the method and materials for its practice. Those aspects are as follows:

- the method for determination,
- a library of modified polypeptides suitable for use in the method,
- 15 a second library of peptide fragments (modified domains) suitable for generating the library of modified polypeptides,
- a third library of DNA or RNA sequences encoding the library of modified polypeptides,
- a fourth library of DNA or RNA sequences encoding the peptide fragments (modified domains) of the second library,
- 20 a library of expression vectors containing the DNA or RNA sequences of the third library,
- a method for synthesis of each of the libraries based upon solid phase peptide synthesis or a combination of solid phase synthesis and recombinant DNA expression
- 25 specific libraries based upon the bHLH transcription factors exemplified.

The method for determining the interaction between a polypeptide and a binding agent is based upon facile, rapid formation of a library of systematically varied polypeptide sequences and the analysis of the entire library without its separation. The method includes the steps of contacting a library of modified polypeptides with a binding agent known to interact with a lead polypeptide, and determining which of the members of the library have bound to the binding

agent. The determination may be carried out by any analytic method that simultaneously analyzes all members. Such methods include mass spectrometry, electrophoresis, high pressure liquid chromatography, two dimensional electrophoresis, gel permeation separation, nuclear magnetic resonance, or
5 infrared spectroscopy.

The library of modified polypeptides is based upon the amino acid sequence of the lead polypeptide. As stated above, the lead polypeptide is known to bind to the binding agent and that binding is the interaction to be studied. Preferably, the lead polypeptide and the library of modified
10 polypeptides have conformationally constrained configurations. The lead polypeptide has an amino acid sequence of at least two parts: a constant region and a selected domain. The constant region may be a contiguous amino acid sequence or may be discontinuous amino acid sequences. The selected domain has the amino acid sequence that is to be studied to determine its interaction with
15 the binding agent. This domain may be a primary binding site, a secondary binding site, an allosteric site, or any site that directly or indirectly participates in an interaction with the binding agent.

The modified polypeptides all have the same constant region which has the same sequence and location that of the lead polypeptide. Each member of
20 the library also has a modified domain that occupies the same location that of the selected domain occupies in the lead polypeptide. Each modified domain, however, has an amino acid sequence that is one or more amino acid unit deletions, substitutions, additions and/or modifications of the amino acid sequence of the selected domain. Together, the group of modified domains is a
25 second library of peptide fragments that represent systematic variation of the amino acid sequence of the selected domain.

The library of peptide fragments is based upon the selected domain of the lead polypeptide. Using the amino acid sequence of the selected domain as a template, the fragments are produced by deleting, substituting, adding or
30 modifying one or more amino acids of the template. Systematic variation is used to produce the library. In this fashion, a systematic study of the interaction of the selected domain with the binding agent can be accomplished.

The libraries may include systematic deletions of amino acid units of the selected domain so as to produce peptide fragments having from one up to the same number of amino acid units as the selected domain.

The libraries may also include systematic substitutions of amino acids
5 such as a conservative or non-conservative substitution of a natural or non-natural amino acid, for example glycine, alanine, serine, leucine, tryptophan, tyrosine or a non-natural amino acid for one or more of the amino acid units of the selected domain. These substitutions may follow the substitution groupings of Kyte and Doolittle (1982). U.S. Pat. No. 6,020,312. Use of glycine may
10 provide a spacer amino acid unit that does not contribute to hydrogen bonding, cationic or anionic interaction, polar interaction, or lipophilic interaction. In this fashion the activity of the unit for which glycine is substituted may be examined. If size of the unit is to be studied, substitution by other amino acid units having larger side chain sizes, such as leucine or tryptophan may be made. If polarity
15 and/or hydrogen bonding are to be studied, substitution by other amino acid units having such characteristics, such as serine, may be made.

The libraries may also include systematic additions to the selected domain so as to determine information about domain size and binding fit. For example, use of mono or multiglycine units can accomplish this purpose.
20 Preferably, the additions may be up to about 10 amino acid units. This size is thought to be the typical binding site size of a peptide fragment. Of course, the additions may include cationic, anionic, hydroxyl or lipophilic amino acid units. These units may provide further information about binding interaction.

The libraries may also include modifications of the selected domain.
25 These modifications are directed to the peptide linkage between amino acid units or to modification of a unit side chain. Such selected units may be modified so that the linkage between them is an ester, thioester, carbonate, allyl or nitro, methoxy phenylmethylamido group. Such selected units may alternatively be modified by alkylation, esterification or acylation of functional groups on the
30 unit side chain. These groups can be selectively cleaved by appropriate reagents and the polypeptide fractions produced will provide a tool for determining the sequence of the corresponding modified domain. It is preferred to provide such modifications when the molecular weights of the modified domains are all the

same or are substantially close such as where glycine is substituted for any of four lysines in a selected domain. The selective cleavages will enable determination of the identity of the modified polypeptide in this instance.

Another aspect of the invention is the DNA or RNA sequences encoding
5 the modified polypeptides and the modified domains. Libraries of the DNA or RNA sequences, recombinant expression vectors containing such libraries and transfected organisms containing libraries of such vectors are also included. Such libraries, of course, provide sequences encoding natural amino acids. Modified polypeptides having non-natural amino acid units or modifications of
10 the selected domain may be made by a combination of expression of DNA or RNA sequences for the constant regions and/or the constant regions and modified domains that contain naturally occurring amino acid units, and solid phase synthesis of the modified domains or portions thereof that contain non-natural amino acid units.

15 The libraries of modified polypeptides, modified domains and DNA or RNA sequences may be produced by solid phase peptide or DNA/RNA synthesis, recombinant expression or a combination thereof. In each instance, solid phase synthesis is employed to produce rapidly the library of sequences presenting the systematic variations. Where the individual members of the
20 polypeptide libraries are 200 amino acid residues in length or less, total chemical synthesis using solid phase techniques is preferred. Polypeptides that are longer than 200 amino acids residues in length can be prepared using a combination of methods.

If modified polypeptides are to be produced, solid phase peptide
25 synthesis is employed to produce the library of modified domains. A synthetic scheme is planned so that all the desired variations are produced by a minimum number of domain syntheses. After each or selected amino acid additions, the product is divided into separate portions. The portions are separately employed to provide the desired deletion, substitution, addition or modification. Then if
30 appropriate, the portions may be recombined to complete the remaining amino acid additions. If not appropriate, the portions are separately reacted through the remaining amino acid sequence. Appropriate amino and carboxy-protecting

groups may be used throughout the solid phase synthesis to provide selectivity and to control the sequential addition of amino acid units.

In similar fashion, libraries of oligonucleotides encoding the library of modified domains may be produced by solid phase nucleotide synthesis.

- 5 Appropriate hydroxyl and phosphate protecting groups may be used throughout the solid phase nucleotide synthesis to provide selectivity and to control the sequential addition of nucleotide units. As mentioned above, the libraries of oligonucleotides provide sequences encoding natural amino acid units.

- 10 The libraries of modified domains or nucleotides encoding the modified domains may be ligated to the constant region of the modified polypeptide or the nucleotide sequence encoding the constant region to form the libraries of modified polypeptides or nucleotide sequences encoding the modified polypeptides. The modified polypeptides may be used as described in the foregoing description of the method of the invention.

- 15 The libraries of nucleotide sequences encoding the modified polypeptides may be inserted into expression vectors such as plasmids, phages or viruses. The vectors may be transfected or infected into eukaryotic or prokaryotic cells such as CHO cells, immortal myeloma cells, E. coli, B. subtilis and the like. The vectors may carry appropriate promoters, introns, and signal regions to provide
20 for expression of the nucleotide libraries. Culturing the recombinant cells may produce the desired libraries of modified polypeptides as extracellular secretions or as intracellular material. The cells may be lysed to obtain the intracellular material.

- 25 Yet another aspect of the invention is application of the method for determination of peptide sequences that will bind a selected binding agent. In this instance, a lead polypeptide is not available. Libraries of modified domains are synthesized based upon the three-dimensional configuration and functional character of the selected binding agent. Typically the modified domains will be no larger than 100 units. A proposed selected domain based upon the
30 functionality and configuration of the binding agent is set out as the template. Systematic variation of this selected domain to produce the modified domains is accomplished as described according to the invention. The library of modified domains is combined with an immobilized version of the selected binding agent

and the method of the invention carried out. A determination of the complexes of modified domains with binding agent according to the invention will provide identification of peptide sequences that will bind to the binding agent.

Reiteration of the application may further refine the identification of peptide
5 sequences that will bind. These peptide sequences may then be incorporated as substitutes for binding sites in proteins such as antibodies, transcription factors and the like. Recombinant methods may be used to produce such proteins.

A further aspect of the invention is the library of modified domains based upon a deletion or glycine substitution for certain amino acid units of a selected
10 domain of a basic helix loop helix (bHLH) transcription factor. Preferably, the selected domain is the loop region. Preferably the bHLH transcription factor is from *Drosophila*. Other preferred libraries include such transcription factor basic domains as the leucine zipper factors (bZIP), the helix loop helix/leucine zipper factors (bHLH-ZIP), , NF-1, RF-X, bHSH, zinc coordinated binding domains,
15 helix turn helix domains, and beta scaffold factors.

Brief Description of the Figures

FIG. 1A shows the amino acid sequence of the bHLH domain of Deadpan (residues 39-102)(SEQ ID NO:3).

20 FIG. 1B shows four libraries containing successive, single amino acid deletions (SAD) in the N-terminal or C-terminal loop region (SEQ ID NO:4 through SEQ ID NO:31).

FIG. 1C shows a MALDI mass spectrum of each SAD library.

FIG. 1D illustrates a schematic and MALDI mass spectrum of the
25 internal amino acid deletion (IAD) library (SEQ ID NO:4 and SEQ ID NO:32 through SEQ ID NO:35).

FIG. 2A shows a MALDI mass spectra of the N SAD-L library before (top) and after DNA affinity selection.

FIG. 2B shows an EMSA of selected elution fractions from a DNA
30 affinity column.

FIG. 2C shows a DNA affinity selection of the IAD peptide library. Ion signals corresponding to WT-Dpn and a mutant missing two amino acids.

FIG. 2D shows MALDI mass spectra of libraries after bHLH affinity selection.

FIG. 3A shows the structure of the Ala-O-Gly linker incorporated into the loop region of Dpn (top).

5 FIG. 3B shows the position of the Ala-O-Gly linker in the loop region sequence (SEQ ID NO:36 through SEQ ID NO:46).

FIG. 3C shows MALDI mass spectra of the library before and after application to the DNA affinity column.

10 FIG. 4A shows a chemical representation of the WT-Dpn side chain (Lys) and the two unnatural amino acid substitutions (Nle and Orn).

FIG. 4B shows a graphical representation of EMSA peptide titrations (26) for WT-Dpn, Dpn Nle 80, and Dpn Orn 80.

FIG. 4C shows the DNA binding specificity of Dpn Nle 80 in comparison to WT-Dpn.

15 FIG. 5 provides a schematic depicting the preparation of modified polypeptide libraries.

Definitions

The term "peptide" means a polymeric compound formed by the condensation of two or more amino acids.

20 The term "polypeptide" means a naturally occurring or synthetic (recombinant or chemical) molecule composed essentially of amino acids, typically linked together by their amino and carboxy groups, and possessing functional fragments that are conformationally constrained so as to allow for specific, selective interaction with other specific molecules. The term

25 "polypeptide" also includes a polypeptide composed of natural and unnatural amino acids and bearing conventional amino protecting groups at the N-terminus or on sidechains (e.g. acetyl or benzyloxycarbonyl), as well as carboxy protecting groups at the C-terminus or on sidechains (e.g. as a (C₁-C₆)alkyl, phenyl, phenethyl, or benzyl ester or amide; or as an -methylbenzyl amide).

30 Other suitable amino and carboxy protecting groups are known to those skilled in the art (See for example, T.W. Greene, T.W.; Wutz, P.G.M. Protecting Groups In Organic Synthesis, Second Edition, 1991, New York, John Wiley & sons, Inc, and references cited therein).

The term "amino acid" includes the residues of natural amino acids and also includes unnatural amino acids. The stereochemistry of amino acids is specified with the D,L system, which is well known to practitioners in the art. Unless otherwise stated peptides of the present invention are composed of amino acids in the L configuration

In keeping with standard polypeptide nomenclature, J. Biol. Chem., 243:3552-59 (1969) and adopted at 37 C.F.R. §1.822(b)(2)), abbreviations for L-amino acid residues are shown in Table 1.

Table 1: Amino Acid Abbreviations

	1-Letter	3-Letter	Amino Acid
	Y	Tyr	tyrosine
	G	Gly	glycine
15	F	Phe	phenylalanine
	M	Met	methionine
	A	Ala	alanine
	S	Ser	serine
	I	Ile	isoleucine
20	L	Leu	leucine
	T	Thr	threonine
	V	Val	valine
	P	Pro	proline
	K	Lys	lysine
25	H	His	histidine
	Q	Gln	glutamine
	E	Glu	glutamic acid
	W	Trp	tryptophan
	R	Arg	arginine
30	D	Asp	aspartic acid
	N	Asn	asparagine
	C	Cys	cysteine

Other amino acids contemplated for use in the present invention include L-alanine, L-arginine, L-aspartic acid, L-asparagine, L-cysteine, L-cysteine, L-glutamic acid, L-glutamine, L-glycine, L-histidine, L-isoleucine, L-leucine, L-lysine, L-methionine, L-phenylalanine, L-proline, L-serine, L-threonine, L-tryptophan, L-tyrosine, L-valine, D-alanine, D-arginine, D-aspartic acid, D-asparagine, D-cysteine, D-cysteine, D-glutamic acid, D-glutamine, D-glycine, D-

histidine, D-isoleucine, D-leucine, D-lysine, D-methionine, D-phenylalanine, D-proline, D-serine, D-threonine, D-tryptophan, D-tyrosine, D-valine, L- α -aminobutyric acid, D- α -aminobutyric acid, L- α -aminobutyric acid, D- γ -aminobutyric acid, L- ϵ -aminocaproic acid, D- ϵ -aminocaproic acid, L-homophenylalanine, D-homophenylalanine, L-alloisoleucine, D-alloisoleucine, L-2-naphthylalanine, D-2-naphthylalanine, L-norvaline, D-norvaline, L-ornithine, D-ornithine, L-pyridyl alanine, D-pyridyl alanine, L-2-thienylalanine, D-2-thienylalanine, L-methyltyrosine, D-methyltyrosine, L-citrulline, D-citrulline, L-homocitrulline, D-homocitrulline, 3-aminomethyl benzoic acid, 4-aminomethyl benzoic acid, diethyl glycine, phosphoserine, phosphothreonine, phosphotyrosine, hydroxyproline, gamma-carboxyglutamate, hippuric acid, octahydroindole-2-carboxylic acid, statine, 1,2,3,4-tetrahydroisoquinoline-3-carboxylic acid, penicillamine, ornithine, citrulline, N-methyl-alanine, para-benzoylphenylalanine, phenylglycine, propargylglycine, sarcosine, and tert-butylglycine.

The term "peptide fragment" means a smaller portion of a polypeptide, wherein the peptide fragment is a binding domain.

The term "selected domain" is used to define a functional fragment of a polypeptide which includes all or part of the molecular elements which effect a specified function such as substrate binding, bactericidal properties, receptor binding, immune stimulation, etc.

The term "constant region," as in the phrase "constant region amino acid sequence," means a region of a given polypeptide wherein the amino acid sequence of the region is not covalently modified by addition or deletion of an amino acid, or by substitution of one amino acid for another.

The term "modified domain," as in the phrase "modified domain amino acid sequence," means a region of a given polypeptide wherein the amino acid sequence follows that of a selective domain but is covalently modified by addition, deletion, substitution, or modification.

The term "lead polypeptide" means a polypeptide known to interact with a binding agent. The lead polypeptide contains constant domain regions and selected domain regions.

The term "linker" means a dipeptide which is formed from two amino acids or amino acid analogues which is substituted at each possible dipeptide position within the selected domain.

5 The term "library" means a large collection of different molecules such as polypeptides or oligonucleotides, with many possible combinations of amino acids or nucleic acids joined together.

The term "solid phase peptide or nucleotide synthesis" means the technique of preparing molecules such as polypeptides and nucleotides in which the polypeptide or nucleotide is anchored to an insoluble support or resin. Solid-
10 phase chemical peptide synthesis methods have been known in the art since the early 1960's (Merrifield, R. B., J. Am. Chem. Soc., 85, 2149-2154 (1963) (See also Stewart, J. M. and Young, J. D., Solid Phase Peptide Synthesis, 2 ed., Pierce Chemical Co., Rockford, Ill., pp. 11-12) and have recently been employed in commercially available laboratory peptide design and synthesis kits (Cambridge
15 Research Biochemicals). Such commercially available laboratory kits have generally utilized the teachings of H. M. Geysen et al, Proc. Natl. Acad. Sci., USA, 81, 3998 (1984).

The term "recombinant expression" means the cellular expression of a nucleotide sequence encoding a modified polypeptide or constant region so as to
20 produce the modified polypeptide or constant region.

The term "vector" means a vehicle to allow insertion, propagation and expression of a gene or nucleotide sequence, and includes a plasmid, cosmid, phage or the like.

The term "host" means any cell that will allow expression of modified
25 polypeptides.

The term "promoter(s)" means regulatory DNA sequences that control transcription of cDNA.

The term "multiple cloning cassette" means a DNA fragment containing unique restriction enzyme cleavage sites for a variety of enzymes allowing
30 insertion of a variety of cDNAs.

The term "primer" referred to herein includes naturally occurring, and modified nucleotides linked together by naturally occurring, and non-naturally occurring oligonucleotide linkages. Primers typically consist of 200 bases or

fewer in length. Preferably, oligonucleotides are 10 to 60 bases in length and most preferably 12, 13, 14, 15, 16, 17, 18, 19, or 20 to 40 bases in length.

Oligonucleotides are usually single stranded. Oligonucleotides can be either sense or antisense oligonucleotides. The term "naturally occurring nucleotides" referred to herein includes deoxyribonucleotides and ribonucleotides.

5 referred to herein includes deoxyribonucleotides and ribonucleotides.

The term "transformation" means incorporation permitting expression of heterologous DNA sequences by a cell.

Detailed Description of the Invention

The present method provides a method for determining the interaction
10 between a lead polypeptide and a binding agent. The lead polypeptide can be an enzyme, DNA binding protein, RNA binding protein, antibody, kinase, G protein, lipoprotein, chemical messenger binding protein, or the like. Suitable lead polypeptides include adrenocorticotrophic hormone, angiotensin I-III, bradykinins, dynorphins, endorphins, enkephalins, gastrin and gastrin-related
15 peptides, glucagon-like polypeptide, bombesins, cholecystokinins, galanin, gastric inhibitory peptides, gastrin-releasing peptide, motilin, neuropeptide Y, pancreastatin, secretin, vasoactive intestinal peptide, growth hormone, growth hormone releasing factor (GRF), luteinizing hormone releasing hormone (LHRH), melanocyte stimulating hormones (MSH), neurotensins, nerve growth
20 factor (NGF), somatostatin, substance P, atrial natriuretic peptide (ANP), corticotropin releasing factors, epidermal growth factor, insulin, thymosin, calcitonin, urotensin, and the like. Other suitable lead polypeptides include fragments of larger proteins, such as tissue plasminogen activator (tPA) and erythropoietin (EPO), and antigenic epitopes derived from infectious organisms,
25 for example, peptides derived from malarial circumsporozoite antigens or chlamydia major outer membrane protein antigens.

The method involves contacting a library of modified polypeptides with a binding agent known to interact with a lead polypeptide to form a library – binding agent mixture. The modified polypeptides have sequences based upon
30 that of the lead polypeptide, which has at least one constant region amino acid sequence and a selected domain amino acid sequence. Each member of the library of modified polypeptides has the same constant region amino acid sequence as the lead polypeptide. Additionally, each member of the library of

modified polypeptides has a modified domain amino acid sequence that is one or more amino acid unit additions, deletions, substitutions, modifications or the like of the selected domain amino acid sequence of the lead polypeptide. The members of the library of modified polypeptides that have bound to the binding agent are then determined.

The present invention incorporates each modified domain of a modified domain library with the constant region to form the modified polypeptide library. Each modified polypeptide containing a modified domain is displayed with in the context of otherwise native protein structures. These modified polypeptides preferably have conformationally constrained configurations. This structural feature means that the modified polypeptides have tertiary structures that resemble the tertiary structures of native proteins. However, the modified polypeptides do not necessarily adopt the same tertiary structure of native proteins. Nevertheless, with the conformationally constrained configurations established within the context of the modified polypeptides, the library of modified domains provides a means for systematic study of remote and distal binding interactions of large, native-like proteins. The modified domains are not free to adopt multiple or changeable configurations as may occur with small peptides of for example 10 units in size.

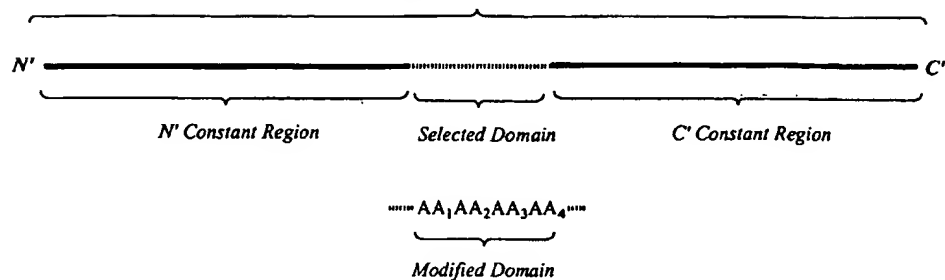
The modified polypeptides with the modified domains are preferably conformationally constrained generally, and in particular, at the modified domain. The size of the modified polypeptides contributes to their conformational constraint. Modified polypeptides will have tertiary structures that resemble the tertiary structures of native proteins. It has been found that these conformationally constrained modified polypeptides show conformational relation to native systems, especially large native systems. Because these conformationally constrained polypeptides resemble the tertiary structure of native systems, the method of the present invention can be applicable to investigation of larger polypeptides, such as proteins having significant conformational character. The present invention also employs analogue units, called linkers, that are easy to manipulate.

A. Modified Polypeptide Libraries

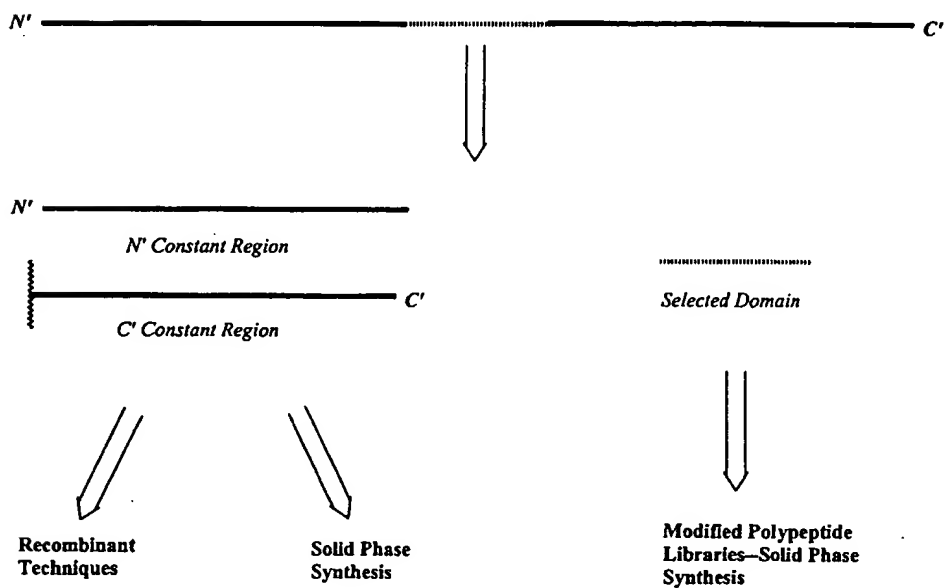
The invention includes a modified polypeptide library. Each member of a modified polypeptide library has at least one constant region amino acid sequence and one selected domain amino acid sequence. The selected domain amino acid sequence may be located at either end of a constant region, or may be positioned between constant regions. One possible representation of a modified polypeptide of the invention is depicted in Scheme 1. Scheme 1 shows the structure of a lead polypeptide with one selected domain amino acid sequence located between two constant region amino acid sequences. In a preferred embodiment of the invention, the lead polypeptide comprises at least two constant region domains (the N' constant region domain and the C' constant region domain) and a selected domain. The selected domain is suitably positioned between the two constant regions, but may also be placed at either the N- or C-terminus of a constant region.

Scheme 1

LEAD POLYPEPTIDE



SYNTHESIS OF LEAD POLYPEPTIDES



In the context of the present invention, a constant region amino acid sequence may be a region that does not interact with binding agents other than to delineate the conformational environment of a selected domain through distal effects. In contrast, a selected domain amino acid sequence may be a region that does interact with a binding agent. That is, the selected region may typically be a binding site or other similar region that is believed to interact with other molecules. The conformational mobility of the selected region amino acid sequence is restricted relative to a linear peptide sequence. The amino acid sequence of a lead polypeptide selected domain is typically no more than 100 amino acid units in length.

The modified domain amino acid sequence, depicted as $[AA_1AA_2AA_3\dots]$ in Scheme 1 is a variant of the selected domain amino acid sequence. The modified domain amino acid sequence contains amino acid additions, deletions, substitutions or modifications of the amino acid sequence of the selected domain. An amino acid addition can be the addition of a natural or non-natural amino acid at a position along the selected domain amino acid sequence. An amino acid deletion can be a deletion of an amino acid or amino acids from the selected domain. An amino acid substitution is the substitution of a natural amino acid present within the specific domain with another natural amino acid, or the substitution of an unnatural amino acid for a natural amino acid. A modification can be a modification of an amino acid within the selected domain by, for instance, alkylation, esterification or acylation of sidechains; or by modification of an amide ($--CONH--$) that connects two adjacent amino acids in the polypeptide chain, by, for instance, replacement with a linkage such as - $NHN(R)CO-$, $-NHB(R)CO-$, $-NHC(RR')CO-$, $-NHC(=CHR)CO-$, $-NHC_6H_4CO-$, $-NHCH_2CHRCO-$, $-NHCHRCH_2CO-$, $-COCH_2-$, $-COS-$, $-CONR-$, $-COO-$, $-CSNH-$, $-CH_2NH-$, $-CH_2CH_2-$, $-CH_2S-$, $-CH_2SO-$, $-CH_2SO_2-$, $-CH(CH_3)S-$, $-CH=CH-$, $-NHCO-$, $-NHCONH-$, $-CONHO-$, and $-C=(CH_2)CH_2-$, or the like.

Amino acid additions and modifications may benefit modified polypeptide identification preferably in situations where the molecular weights of the various members of the modified polypeptide library are not significantly different. An addition may be of an amino acid sequence that is unique to the modified polypeptide and is enzymatically cleavable. Such groups and

enzymatic reactions are known in the art, for example, U.S. Patent No. 5,595,887 and *Enzyme Structure and Function*, 2nd Ed., A. Fersht, Freeman pub., New York, 1985. A modification may be of a backbone group that is readily cleavable by mild chemical methods. The addition or modification is strategically positioned relative to the modified domain and/or its other variations so that when cleaved, the resulting fragment(s) will have distinctly different molecular weights. In this fashion, the simultaneous identification of library members based upon molecular weight or other properties found in a cleaved portion of the modified polypeptide is facily determined.

10 A preferred embodiment of the present invention is the second library of modified domains, wherein each member of the modified domain library has an amino acid sequence that is one or more amino acid unit additions, deletions, substitutions, modifications or combinations thereof of the sequence of the selected domain. Each member of the modified polypeptide domain library has an altered amino acid sequence in its selected domain relative to that of a lead polypeptide.

The modified domains preferably have amino acid sequences of up to 100, preferably up to 40 amino acid units. The modified domain preferably may contain at least one cleavable non-amide linkage joining at least two of the units of the domain. This cleavable, non-amide linkage can be an ester, thioester, carbonate, allyl, or nitro, methoxy phenyl amide linkage. The amino acid units within the sequences can be randomly varied. The selected domain and modified domain of a modified polypeptide library are preferably no more than 100 amino acid units in length. Preferably, the selected domain and the modified domains of a library are no more than 50 amino acid units in length.

25 In one embodiment of the present invention, a library of modified polypeptide domains can be formed by substituting a second library of peptide fragments for the selected domain of the lead polypeptide. Each member of the second library is covalently bound to the constant region to form each member of the library of modified polypeptides. Preferably, all members of the second library are simultaneously bound to constant regions to form the library of modified polypeptides.

The amino acid sequence of each member of the second library is one or more amino acid unit additions, deletions, substitutions, modifications or combinations thereof of the amino acid sequence of the selected domain. Each peptide fragment of the second library can have the sequence of the selected domain except that one or more amino acid units are deleted from the selected domain sequence to produce each peptide fragment. The peptide fragments of the second library can all have the same number of peptide units as the selected domain, and one or more amino acid units such as conservative or non-conservative substitutions including, but not limited to, glycine, alanine, leucine, tyrosine, tryptophan, or serine units or combinations thereof, as well as unnatural amino acids are substituted for one or more selected amino acid unit of the selected domain to form each peptide fragment.

The second library also may be a final product having a final amino acid sequence and a group of intermediates having amino acid sequences that are one or more deletions from the final amino acid sequence. The group of intermediates having amino acid sequences can have one or more glycine, alanine or serine units substituted for one of more selected amino acid units of the final amino acid sequence.

The invention includes a third library of DNA or RNA oligonucleotide sequences that encode a library of modified polypeptides where the modified polypeptides contain natural amino acid units. The library of DNA or RNA nucleotide sequences can be used to recombinantly express the modified polypeptide libraries, preferably large polypeptides. Modified polypeptides having non-natural amino acid units or linkages can be expressed using DNA or RNA expression and semisynthetic techniques, such as those found in Muir et. al., *Proc. Nat'l Acad. Sci. USA* 95 6705-6710 (1998). Accordingly, chemosynthetic peptides representing a modified domain library, can be added to a constant region or regions produced by recombinant expression using for example a thioester-cysteine leaving group reaction. The thioester generated as the C terminus group on a member of the modified domain library (produced by solid phase synthesis) is intercepted (reacted) with an N-terminal cysteine on the N terminus of a constant region or regions (produced by recombinant expression).

A fourth library of DNA or RNA oligonucleotide sequences that encodes the second library of modified domains having natural amino acid units is included also. The fourth library is produced by solid phase synthesis. The members of the fourth library can be ligated to the DNA or RNA sequences encoding the constant region of the modified polypeptides.

B. Preparation of Modified Polypeptide and Nucleotide Libraries

1. Preparation of Modified Polypeptide and Nucleotide Libraries

a. Constant Regions of Lead Polypeptide

In one embodiment of the invention, a constant region of a lead polypeptide or the corresponding nucleotide sequence encoding the constant region can respectively be produced by a suitable chemical technique or by cloning and amplification techniques discussed below. Provided that the sequence is of low or moderate length, the constant region nucleotide sequences can also be produced by solid phase techniques as discussed below for oligonucleotides encoding the modified domains. Sequential chemical peptide and oligonucleotide syntheses are well established, widely used procedures for producing peptides and oligonucleotides, such as those up to and over about 200 residues (peptides) and up to and over about 600 residues (oligonucleotides). For peptides, the chemistry involves the specific coupling of the amino terminus of a carboxyl-blocked peptide to the activated carboxyl group of an amino-blocked amino acid. For oligonucleotides, the chemistry involves the specific coupling of the 5'-hydroxyl group of a 3'-blocked nucleotide to an activated 3'-hydroxyl group of a 5'-blocked nucleotide. A description of solid phase synthesis can be found in Abelson, John M; Simon, Melvin I. Methods in Enzymology: Solid-Phase Peptide Synthesis (New York: Academic) (1997).

In their most commonly used forms, developed primarily by Merrifield, J. Amer. Chem. Soc., 85, 2149 (1963) and Beaucage, S. L. and Caruthers, M. H., Tet. Lett., 22, 1859-1862 (1981); Beaucage, S. L. and Caruthers, M. H., J. Amer. Chem. Soc., 24, 3184-3191 (1981), these syntheses are accomplished with the peptide or oligonucleotide immobilized on a solid support. An extremely large number of peptides or oligonucleotides can be produced by this methodology. The physical and chemical properties of the peptide or oligonucleotide products

will vary greatly depending on size and composition of the respective amino acids or nucleotides composing these products. Consequently, it is typical to tailor the synthetic techniques to fit the specific product at hand.

In the method of immobilized peptide synthesis, the carboxyl terminal amino acid is bound to a polyvinyl benzene or other suitable insoluble resin. The second amino acid to be added possesses blocking groups on its amino moiety and any side chain reactive groups so that only its carboxyl moiety can react. This carboxyl group is activated with a carbodiimide or other activating agent and then allowed to couple to the immobilized amino acid. After removal of the amino blocking groups the cycle is repeated for each amino acid in the sequence.

b. Modified Domain Libraries

The second library of modified domain libraries can be produced by a suitable technique such as solid phase peptide synthesis. When solid phase synthesis techniques are used to prepare the library of modified domains, amino acid units are sequentially reacted together by chemical synthetic techniques. After addition of each amino acid unit, a portion of the resulting intermediate is isolated, and the remaining portion is used as the starting material for addition of a further amino acid unit until the final product is produced. After addition of each amino acid unit, a first portion of the resulting intermediate can be isolated. A conservative or non-conservative amino acid substitution unit can be reacted with the first portion to form a second intermediate. The corresponding amino acid unit of the final amino acid sequence can be added to the remaining portion to form a third intermediate. The additional amino acid units of the final amino acid sequence can be added to form a final sequence for the second intermediate.

The remaining portion is used as the starting material for addition of a further amino acid unit until the final product is produced. The remaining portion can be used as the starting material for addition of a further amino acid unit, and the formation of a first portion and remaining portion is repeated after each amino acid unit addition until the final product and library are produced.

Thus, a library of modified domains containing single amino acid substitutions can be prepared as follows and as depicted in Figure 5. Two manual solid phase peptide synthesis (SPPS) reaction vessels, A and B, and a

small fritted funnel, 1, are used to manipulate peptide resin. The synthesis begins with ten units of peptide-resin in vessel A. After deprotection of the α -amino group, one unit of peptide-resin is removed from A and added to 1. The first amino acid is then coupled to the nine units of peptide-resin in A and the analogue moiety to the one unit peptide-resin sample in 1. After the coupling step, the analogue-modified peptide-resin from 1 is transferred to B. To initiate the next cycle of synthesis, the peptide-resins in vessels A and B are deprotected. Another unit of peptide-resin is removed from A and transferred to the now empty 1. The next amino acid in the sequence of the parent peptide is added in activated form to both A and B, while the substitution amino acid is reacted with the new peptide-resin sample in 1. After completion of this cycle, the modified peptide-resin in 1 is added to B. The synthesis continues in this manner for the requisite ten cycles.

Throughout the synthesis, vessel A contains only unmodified peptide-resin. Vessel B contains all single-site modified peptide-resins and vessel 1 contains the current sample of peptide-resin which is being modified. All chemical steps carried out in vessels A and B are identical, adding the amino acids of the unmodified sequence. At the end of 10 cycles, all the resin in vessel A has been transferred into vessel B which now contains the desired array of peptide analogues in resin-bound form.

A dipeptide linker can be incorporated into a library of modified polypeptides using a similar procedure. However, since the analogue moiety is preferably incorporated as a dipeptide, a modification can be made to the synthetic procedure outlined above. In order to keep the synthetic operations being performed on the peptides in vessels A and B in register, the sample being derivatized in 1 is held out for two cycles before transfer to vessel B. To accommodate this modification, a second auxiliary funnel is added. The peptide-resin sample from vessel A is added to a funnel in position 1, where the linker analogue coupling is initiated. After one cycle, the funnel is moved to the new funnel position, where the dipeptide coupling continues during a second cycle of chain elongation in vessels A and B. The analogue-containing sample of peptide-resin is then washed with DMF (dimethylformamide) and transferred

to vessel B. The dipeptide linker is substituted for consecutive dipeptide sequences spanning a region of a selected domain.

c. Oligonucleotide Libraries Encoding Modified Domains

- Synthesis of oligonucleotide libraries that encode addition, substitution, or deletion modified domains of naturally occurring amino acid units can be accomplished using both solution phase and solid phase methods. A general review of solid-phase versus solution-phase oligonucleotide synthesis is given in the background section of Urdea et al. U.S. Pat. No. 4,517,338, entitled "Multiple Reactor System And Method For Oligonucleotide Synthesis."
- Oligonucleotide synthesis via solution phase can be accomplished with several coupling mechanisms. One such solution phase preparation utilizes phosphorus triesters. Yau, E. K. et al., *Tetrahedron Letters*, 1990, 31, 1953, report the use of phosphorous triesters to prepare thymidine dinucleoside and thymidine dinucleotide phosphorodithioates. Further details of methods useful for preparing oligonucleotides may be found in Sekine, M. et al., *J. Org. Chem.*, 1979, 44, 2325; Dahl, O., *Sulfer Reports*, 1991, 11, 167-192; Kresse, J. et al., *Nucleic Acids Res.*, 1975, 2, 1-9; Eckstein, F., *Ann. Rev. Biochem.*, 1985, 54, 367-402; and Yau, E. K. U.S. Pat. No. 5,210,264.

- The current method of choice for the preparation of oligonucleotides encoding naturally-occurring amino acids, is via solid-phase synthesis. Solid-phase synthesis involves the attachment of a nucleotide to a solid support, such as a polymer support, and the addition of a second nucleotide onto the support-bound nucleotide. Further nucleotides are added, thus forming an oligonucleotide which is bound to a solid support. The oligonucleotide can then be cleaved from the solid support when synthesis of the desired length and sequence of oligonucleotide is achieved.

- As indicated, solid-phase synthesis relies on sequential addition of nucleotides to one end of a growing oligonucleotide chain. Typically, a first nucleotide, having protecting groups on any exocyclic amine functionalities present, is attached to an appropriate solid support. In general, the oligonucleotide synthetic procedure follows the well-established 3'-phosphoramidite schemes devised by Caruthers. The 3' terminal base of the desired oligonucleotide is immobilized on an insoluble carrier. The nucleotide

base to be added is blocked at the 5' hydroxyl and activated at the 3' hydroxyl so as to cause coupling with the immobilized nucleotide base. Deblocking of the new immobilized nucleotide compound and repetition of the cycle will produce the desired final oligonucleotide.

5 **2. Preparation of Vectors and Containing DNA or RNA
Sequences For Modified Polypeptide Libraries**

Using the DNA, RNA or cDNA sequence encoding the lead polypeptide, "polymerase chain reaction" or "PCR" can be used to amplify the constant regions. PCR refers to a procedure or technique in which amounts of a
10 preselected fragment of nucleic acid, RNA and/or DNA, are amplified as described in U.S. Patent No. 4,683,195. Generally, sequence information from the ends of the region of interest or beyond is employed to design oligonucleotide primers comprising at least 7-8 nucleotides. These primers will be identical or similar in sequence to opposite strands of the template to be
15 amplified. The primers may also optionally contain sequences encoding restriction endonuclease sites to facilitate cloning the PCR product into a suitable vector. PCR can be used to amplify specific RNA sequences, specific DNA sequences from total genomic DNA, and cDNA transcribed from total cellular RNA, bacteriophage or plasmid sequences, and the like. See generally Mullis et
20 al., Cold Spring Harbor Symp. Quant. Biol., 51, 263 (1987); Erlich, ed., PCR Technology, (Stockton Press, New York, 1989).

Primers are made to correspond to nucleotide sequences of the lead polypeptide. One primer is prepared which is predicted to anneal to the antisense strand, and another primer prepared which is predicted to anneal to the
25 sense strand, of a DNA molecule/polynucleotide which encodes a constant region polypeptide, either the N' constant region or the C' constant region.

The products of each PCR reaction are separated via an agarose gel and all consistently amplified products are gel-purified and are then either directly ligated to the oligonucleotide sequence for the modified domain, as described in
30 section D, entitled "Generating Polynucleotide Sequences for a Library of Modified Polypeptides and then cloned by well known recombinant techniques into a suitable expression vector (as described below), or the products of the PCR reaction are cloned directly into a suitable vector, such as a known plasmid

vector so that expression of the constant region can be obtained. The resultant PCR products or plasmids are subjected to restriction endonuclease and dideoxy sequencing of the double-stranded DNAs.

To prepare expression vectors for transformation herein, the recombinant
5 or selected DNA sequence or segment containing either the N' or C' constant region of the lead polypeptide or the oligonucleotide product obtained from Section D, may be circular or linear, double-stranded or single-stranded. Generally, the DNA sequence or segment is in the form of chimeric DNA, such as plasmid DNA, that can also contain coding regions flanked by control
10 sequences which promote the expression of the selected DNA present in the resultant cell line.

As used herein, "chimeric" means that a vector comprises DNA from at least two different species, or comprises DNA from the same species, which is linked or associated in a manner which does not occur in the "native" or wild
15 type of the species.

"Control sequences" is defined to mean DNA sequences necessary for the expression of an operably linked coding sequence in a particular host organism. The control sequences that are suitable for prokaryotic cells, for example, include a promoter, and optionally an operator sequence, and a
20 ribosome binding site. Eukaryotic cells are known to utilize promoters (such as the CMV promoter, as well as the SV40 late promoter and retroviral LTRs (long terminal repeat elements)), although many other promoter elements well known in the art may be employed in the practice of invention), polyadenylation signals, and enhancers.

25 Most genes have regions of DNA sequence that are known as promoters and which regulate gene expression. Promoter regions are typically found in the flanking DNA sequence upstream from the coding sequence in both prokaryotic and eukaryotic cells. A promoter sequence provides for regulation of transcription of the downstream gene sequence and typically includes from about
30 50 to about 2,000 nucleotide base pairs. Promoter sequences also contain regulatory sequences such as enhancer sequences that can influence the level of gene expression. Some isolated promoter sequences can provide for gene expression of heterologous genes, that is a gene different from the native or

homologous gene. Promoter sequences are also known to be strong or weak or inducible. A strong promoter provides for a high level of gene expression, whereas a weak promoter provides for a very low level of gene expression. An isolated promoter sequence that is a strong promoter for heterologous genes is
5 advantageous because it provides for a sufficient level of gene expression to allow for easy detection and selection of transformed cells and provides for a high level of gene expression when desired.

The polynucleotide encoding the constant region of the modified polypeptide or the oligonucleotide product obtained from Section D of interest
10 can be combined with a promoter by standard methods as described in Sambrook cited *supra*. Briefly, a plasmid containing a promoter can be constructed or obtained from a wide variety of commercial vendors, such as the Clontech Lab in Palo Alto, CA. Typically these plasmids are constructed to provide for multiple cloning sites having specificity for different restriction enzymes downstream
15 from the promoter. The constant region polynucleotide or the oligonucleotide product obtained from Section 1c can be subcloned downstream from the promoter using restriction enzymes to ensure that the coding region is inserted in proper orientation with respect to the promoter so that the coding region can be expressed.

20 Other elements functional in the host cells, such as introns, enhancers, polyadenylation sequences and the like, may also be a part of the DNA. Such elements may or may not be necessary for the function of the DNA, but may provide improved expression of the DNA by affecting transcription, stability of the mRNA, or the like. Such elements may be included in the DNA as desired to
25 obtain the optimal performance of the transforming DNA in the cell.

Plasmid vectors included additional DNA sequences that provide for easy selection, amplification and transformation of the expression cassette in prokaryotic and eukaryotic cells. The additional DNA sequences include origins of replication to provide for autonomous replication of the vector, selectable
30 marker genes, preferably encoding antibiotic resistance, unique multiple cloning sites providing for multiple sites to insert DNA sequences or genes encoded in the expression cassette, and sequences that enhance transformation of

prokaryotic and eukaryotic cells. The preferred vectors of the invention are plasmid vectors.

Furthermore, the vector can also optionally include 5' and 3' nontranslated regulatory DNA sequences. The 3' nontranslated regulatory DNA sequence preferably includes from about 300 to 1,000 nucleotide base pairs and contains transcriptional and translational termination sequences. The 3' nontranslated regulatory sequences can be operably linked to the 3' terminus of a coding region by standard methods.

"Operably linked" is defined to mean that the nucleic acids are placed in a functional relationship with another nucleic acid sequence. For example, DNA for a presequence or secretory leader is operably linked to DNA for a peptide or polypeptide if it is expressed as a preprotein that participates in the secretion of the peptide or polypeptide; a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of the sequence; or a ribosome binding site is operably linked to a coding sequence if it is positioned so as to facilitate translation. Generally, "operably linked" means that the DNA sequences being linked are contiguous and, in the case of a secretory leader, contiguous and in reading phase. However, enhancers do not have to be contiguous. Linking is accomplished by ligation at convenient restriction sites. If such sites do not exist, the synthetic oligonucleotide adaptors or linkers are used in accord with conventional practice.

The general methods for constructing recombinant DNA which can transform target cells are well known to those skilled in the art, and the same compositions and methods of construction may be utilized to produce the DNA useful herein. For example, J. Sambrook et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, NY (1989), provides suitable methods of construction.

Expression vectors comprising genes for the constant regions or for the modified polypeptides can be readily introduced into the host cells, e.g., mammalian, bacterial, yeast or insect cells by transfection carried out by any procedure useful for the introduction into a particular cell, e.g., physical or biological methods, to yield a transformed cell expressing the DNA molecules of the present invention.

Physical methods to introduce a DNA into a host cell include calcium phosphate precipitation, lipofection, particle bombardment, microinjection, electroporation, and the like. Biological methods to introduce the DNA of interest into a host cell include the use of DNA and RNA viral vectors. Other
5 viral vectors can be derived from poxviruses, herpes simplex virus I, adenoviruses and ADENO-associated viruses, and the like.

As used herein, the term "cell line" or "host cell" is intended to include well-characterized homogenous, biologically pure populations of cells. These cells may be eukaryotic cells that are neoplastic or which have been
10 "immortalized" *in vitro* by methods known in the art, as well as primary cells, or prokaryotic cells. Additionally, cell lines or host cells which also may be employed include plant, insect, yeast, fungal or bacterial sources.

If the expressed constant region or lead polypeptide was operably linked to a secretory leader, and thus the protein is secreted into the medium, the
15 medium can be recovered and the expressed protein purified therefrom by techniques well known in the art. If the constant region or modified polypeptide is produced intracellularly, the cells must first be lysed. The polypeptide is then recovered from the cell lysate by techniques well known in the art. Additionally, to aid in purification, the modified polypeptide or constant region may also
20 optionally be operably linked to a marker sequence which facilitates purification of the fused polypeptide, for example, the marker sequence can be a hexahistidine (His-tag) peptide, as provided in the pQF vector (Qiagen, Inc.) and described in Gentz et al., Proc. Natl. Acad. Sci. USA (1989) 86:821-824.

The isolated constant region polypeptides are then ligated, as
25 demonstrated in section C entitled "Joining the Constant Regions of the Lead Polypeptide to the Modified Polypeptides of the Modified Polypeptide Libraries", to the modified polypeptide domain library of section b, entitled "Modified Polypeptide Domain Libraries." Suitable ligation techniques include the method of Muir, et. al., in *Proc. Nat'l Acad. Sci. USA* 95: 6705-6710 (1998).
30 *In vitro* transcriptional, translational and folding techniques may also be employed.

C. Joining the Constant Regions to the Modified Domains To Form the Modified Polypeptide Libraries

Instead of synthesizing single polypeptides according to conventional, linear synthesis techniques, the present invention allows for the synthesis of a library of modified polypeptides containing modified domains in a single total chemical synthesis. The modified polypeptides contained in this library differ from each other only by the position in which a defined covalent modification is located within the modified domain of the polypeptide.

Libraries of modified polypeptides can be prepared by combining constant region or regions with each modified domain from the modified polypeptide domain library (cf. Scheme 1). For example, modified polypeptides can be prepared starting from the N terminus or the C terminus, using modified domain libraries with either free N- or C- termini. Typically, the modified domains will be immobilized to a solid support as a result of their solid phase syntheses. The solid support can be used to advantage in the subsequent joining process.

For synthesis of a modified polypeptide starting in the direction of carboxy to amino terminus, a free amino terminus on the C-constant region is required that can be conveniently blocked and deblocked as needed. A preferred amino terminus protecting group is a fluoromethoxycarbonyl group (Fmoc). Fmoc blocked amino termini are deprotected with (DBU) in dichloromethane (DCM) as is well known for polypeptide synthesis. Modified domain libraries that are connected to a solid support at the N terminus are protected at the carboxyl terminus with pentafluorophenyl ester (Opfp). To perform the joining reaction, the C-constant region protected at the C-terminus with Opfp, the deprotected, immobilized modified domains, dimethylformamide (DMF) and hydroxy-benzotriazole (HOBt) are combined as is well known for peptide synthesis. The resulting intermediate incorporates the C-constant region and the modified domains that are connected to a solid support at the N terminus. To complete the preparation of the modified polypeptide, the N-constant region protected at its N-terminus is added to the intermediate after its cleavage from the solid support. The intermediate is cleaved from the solid support and the carboxy terminus of the intermediate remains protected as the Opfp ester. The

intermediate is then allowed to react with the N terminus constant region protected at the N terminus. The library of modified polypeptides can be prepared starting from the N to C terminus using a similar series of transformations.

- 5 An alternative approach for the synthesis of polypeptides that are larger than 100 amino acid units is found in Muir et. al. (1998). In this approach, small synthetic sequences are ligated to much larger recombinant protein fragments using thioester-intein chemistry. In intein chemistry, a polypeptide undergoes an intramolecular rearrangement resulting in the extrusion of an internal sequence
10 (intein) and the joining of the lateral sequences (exteins).

D. Generating Polynucleotide Sequences for a Library of Modified Polypeptides

- In general, the polynucleotide synthetic procedure for joining nucleotide sequences encoding the modified domains and constant regions is strategically
15 the same as for the synthesis of the oligonucleotides discussed above and follows the well-established 3'-phosphoramidite schemes devised by Caruthers. The 3' terminal bases of the oligonucleotide encoding the modified domains are immobilized on an insoluble carrier. The polynucleotide encoding the 3' constant region (C-constant region polynucleotide) is protected at the 5'
20 hydroxyl and activated at the 3' hydroxyl so as to cause coupling with the immobilized oligonucleotides. The polynucleotide encoding the N-constant region can then be attached to the resulting polynucleotide fragment.

E. Binding Studies of Polypeptides Containing Modified Selected Domains

- 25 The invention also includes a method for identifying a polypeptide that binds to a selected agent. The selected binding agent may be an antigen, a substrate, a carbohydrate, a small molecule or the like. The binding agent can be a substrate, DNA sequence, RNA sequence, antigen, antagonist, carbohydrate, lipid, phospholipid, nucleic acid, agonist, inhibitor, protein binding agent, or
30 receptor activator or any other substance that selective binds to a protein. Typically, the binding agent is immobilized by a suitable method such as by being bound to a solid support. The solid support may be any suitable solid support known in the art. For example, the binding agent may be bound to solid

support materials such as microspheres, sephadex, or agarose. According to the method, a library of modified polypeptides can be contacted with a selected binding agent to form a group-binding agent mixture. The individual modified polypeptides that have bound to the binding agent can then be determined.

5 In one embodiment of the invention, the determining step includes contacting a modified polypeptide with a binding agent or the like to form a modified polypeptide-binding agent complex. As a result, modified polypeptides that are not bound to the binding agent can be separated from modified polypeptides that are bound to the binding agent. Techniques that are
10 suitable for determining the modified polypeptides that are bound to the binding agent are known to the art and include techniques such as mass spectrometry. In a specific embodiment of the present invention, the mass spectrometry technique known as matrix assisted laser desorption ionization mass spectrometry can be used to determine the molecular weights of the modified polypeptides bound to
15 the binding agent.

In order determine which members of the modified polypeptide library specifically bind with a binding agent, the library is subjected to binding agent affinity column chromatography. Affinity column chromatography is based on the ability of members of the polypeptide library to reversibly bind to the
20 binding agent. Separation by agent binding can be accomplished by the various affinity methods known in the art. In this method, the agents are immobilized on an inert matrix, such as agarose, polyacrylamide beads, cellulose or other media. Depending on the library of modified polypeptides which is being purified, the immobilized agents may be small molecules such as heterocycles, carbocycles,
25 linear and branched compounds, biological small molecules such as biotin, peptides such as oxytocin, vasopressin, antigens and double- or single-stranded DNA, double- or single-stranded RNA, or other types, lengths, structures or combination of nucleic acids, such as tRNA, Z-DNA, supercoiled DNA, ultraviolet-irradiated DNA or DNA modified by other agents as well as those
30 listed above.

The binding agents may be attached to the solid phase matrix by a variety of methods, including covalent attachment of the agent through hydrogels, carbogels, thiols, carbonyls, amines or by absorbing the agents to a matrix such

as cellulose, which closely binds the agent. For example, the preferred immobilization method for DNA is to use cyanogen-bromide activated Sepharose and to bind the nucleic acids to the activated Sepharose covalently. Alternatively, single-stranded DNA covalently bound to agarose can be
5 purchased commercially from Bethesda Research Labs, Gaithersburg, Md. (Catalog No. 5906SA).

The library of modified polypeptides can be applied to the binding agent in a solution which should satisfy the following criteria: 1) the solution should permit reversible binding of the modified polypeptides to the binding agent,
10 2) the solution should reduce non-specific binding of contaminating proteins to the binding agent, and 3) the solution should not cause damage to the binding agent or modified polypeptide. In general, a neutral buffered solution with physiological saline and 1 mM EDTA will satisfy these criteria.

The bound modified polypeptides from the modified polypeptide library
15 can be eluted from the binding agent affinity column with an eluant gradient which removes the modified polypeptide from the binding agent at a characteristic condition and concentrates the enzyme by the focusing effect of the gradient. A gradient of NaCl up to 1.0 M will in general be sufficient to reverse the binding of most modified polypeptides that are electrostatically
20 bound to binding agents. In appropriate cases, the gradient may be one of another salt, increasing or decreasing pH, temperature, voltage or detergent, or, if desired, a competing ligand may be introduced to replace the agent binding. Other eluants such as denaturants (guanidine, urea, ethanolic solutions) chelators or chaotropic agents may be appropriate depending upon the nature of the
25 binding interaction between the modified polypeptide and the binding agent.

The modified polypeptides from the modified polypeptide libraries that are bound to the affinity column can then be analyzed and identified. The purpose of the analysis is to identify the modified polypeptides that exhibit binding activity. Many techniques are available for analysis of the affinity
30 bound modified polypeptides, including nuclear magnetic resonance spectroscopy, infrared spectroscopy, mass spectroscopy as well as others known in the art. The use of mass spectrometry to analyze the modified polypeptide libraries of the present invention is analogous to the use of gel electrophoresis to

separate nucleotides by length during DNA sequencing and analysis (See Pan et al.; Science. 1991, 254,1361-1364; Hayashibara et al. J. Am. Chem. Soc. 1991, 113, 5104-5106).

Mass spectroscopy useful for analyzing the modified polypeptides of the present invention includes ionization/desorption techniques known as electrospray/ion spray (ES) and matrix-assisted laser desorption/ionization (MALDI). ES mass spectrometry was introduced by Fenn et al. (J. Phys. Chem. 88, 4451-59 (1984); PCT Application No. WO 90/14148) and applications are summarized in recent review articles (R. D. Smith et al., Anal. Chem. 62, 882-89 (1990) and B. Ardrey, Electrospray Mass Spectrometry, Spectroscopy Europe, 4, 10-18 (1992)). The molecular weights of a tetradecanucleotide (Covey et al. "The Determination of Protein, Oligonucleotide and Peptide Molecular Weights by Ion spray Mass Spectrometry," Rapid Communications in Mass Spectrometry, 2, 249-256 (1988)), and of a 21-mer (Methods in Enzymology, 193, "Mass Spectrometry" (McCloskey, editor), p. 425, 1990, Academic Press, New York) have been published. As a mass analyzer, a quadrupole is most frequently used. The determination of molecular weights in femtomole amounts of sample is very accurate due to the presence of multiple ion peaks which all could be used for the mass calculation.

MALDI mass spectrometry, in contrast, can be particularly attractive when a time-of-flight (TOF) configuration is used as a mass analyzer. The MALDI-TOF mass spectrometry was introduced by Hillenkamp et al. ("Matrix Assisted UV-Laser Desorption/Ionization: A New Approach to Mass Spectrometry of Large Biomolecules," Biological Mass Spectrometry (Burlingame and McCloskey, editors), Elsevier Science Publishers, Amsterdam, pp. 49-60, 1990). Since, in most cases, no multiple molecular ion peaks are produced with this technique, the mass spectra, in principle, look simpler compared to ES mass spectrometry.

Japanese Patent No. 59-131909 describes an instrument, which detects nucleic acid fragments separated either by electrophoresis, liquid chromatography or high speed gel filtration. Mass spectrometric detection is achieved by incorporating into the nucleic acids, atoms which normally do not occur in DNA such as S, Br, I or Ag, Au, Pt, Os, Hg.

Amenable mass spectrometric formats for use in the invention include the ionization (I) techniques such as matrix-assisted laser desorption (MALDI), continuous or pulsed electrospray (ESI) and related methods (e. g. Ionspray, Thermospray), or massive cluster impact (MCI); these ion sources can be matched with detection formats including linear or reflector time-of-flight (TOF), single or multiple quadrupole, single or multiple magnetic sector, Fourier transform ion cyclotron resonance (FTICR), ion trap, or combinations of these to give a hybrid detector (e. g. ion trap--time-of-flight). For ionization, numerous matrix/wavelength combinations (MALDI) or solvent combinations (ESI) can be employed. The high resolution and sensitivity (< 1 pmol/component; Chait et al. Science 1992, 257, 1885-1894) of MALDI mass spectrometry allows the characterization of even small quantities of the entire modified polypeptide library.

The invention will now be illustrated by the following non-limiting Example involving basic helix-loop helix transcription factors.

Example

Material and Methods

Combinatorial Solid Phase Peptide Synthesis. Dpn was manually synthesized using stepwise solid phase peptides synthesis (SPPS) methods, according to published in situ neutralization Boc chemistry protocols. Schnölzer, M., Alewood, P., Jones, A., Alewood, D., and Kent, S. B. H. (1992) *Int J Pept Protein Res* 40, 180-193. 4-methylbenzhydramine polystyrene resin was functionalized with residues comprising helix 2 (83-102), and then split in half for the generation of N- and C-terminal libraries. N-terminal deletions in the loop region sequence were easily introduced to one half of helix 2 resin, by transferring equimolar portions of resin after each amino acid coupling step to a separate vessel where no amino acid coupling took place.

DNA Affinity Column. A Dpn-specific DNA affinity column was prepared using complimentary oligonucleotides containing a Dpn recognition sequence: 5'-CGTACGCCGGCCACGCGACAGGTCC-3' (SEQ ID NO:1) (top strand shown, where the underlined sequence is the Dpn binding site). Kadonaga, J. T. and Tjian, R. (1986) *Proc Natl Acad Sci USA* 83, 5889-5893.

The loading capacity of the column was determined to be 2 nmole/100 μ l of resin using a WT-Dpn standard.

DNA Affinity Selection. The following buffer was used in DNA affinity selection experiments: 20 mM Hepes, 1 mM EDTA, 5% glycerol, pH 7.6.

- 5 Initial binding was carried out using buffer containing 100 mM KCl, and elution steps contained increasing KCl concentrations, as indicated in the text and figure captions. Controls were performed to validate that increasing ionic strength competes away weakly bound peptides and selects for high affinity peptides. Equimolar amounts of three Dpn bHLH peptides (WT-Dpn, Dpn(desPA 75, 76),
10 and Dpn(desDPAR 74-77)) with a range of binding affinities (K_d s of 2.6 nM, 4.4 nM, and 44 nM, respectively, for the Dpn site oligonucleotide as determined by EMSA) were pooled and subjected to DNA affinity selection. MALDI-MS analysis of eluted fractions reflected the individual activity of each peptide, i.e. weaker binding peptides eluted at lower ionic strength.

- 15 *MALDI Mass Spectrometry.* Each crude synthetic library was dissolved in 50% acetonitrile, 0.1% trifluoroacetic acid (TFA) to a concentration of 1-5 μ M. A 2 μ l aliquot was mixed with an equal volume of saturated matrix solution (α -cyano-4-hydroxycinnamic acid in 50% acetonitrile, 0.1 % TFA in water), and 1 μ l of the resulting mixture was placed on a MALDI plate and
20 quickly dried with a heat gun. MALDI mass spectra were collected using a Thermo BioAnalysis DYNAMO mass analyzer with delayed extraction and calibrated with an external standard. Typically, the ion signals generated from 50 laser pulses were summed to give a single mass spectrum. Only signals for the singly charged molecules of bHLH mutants are detected.

- 25 *bHLH Affinity Column.* The polypeptide H-Cys-Ahx-[WT-Dpn (39-102)] (where Ahx is amino hexanoic acid) was synthesized and purified using the general procedures, and reacted with pre-swollen Sulfolink resin (Pierce) under conditions suggested by the manufacturer. Winston, R. L., Millar, D. P., Gottesfeld, J. M., and Kent, S. B. H. (1999) *Biochemistry* 38, 5138-
30 5146. Tris-2-carboxyethyl phosphine/HCl (10 mM, pH 8.3) was added to the coupling reaction to prevent peptide disulfide formation. The functional substitution of the column was determined by Bradford assay to be approximately 200 μ M.

bHLH Affinity Selection. WT-Dpn (1 μ M in the following assay buffer: 100 mM KCl, 1 mM EDTA, 20 mM Hepes, 5% glycerol, pH 7.6) and 83 μ g/ml BSA were incubated with 200 μ l of packed bHLH column resin for 30 minutes with gentle agitation. After washing with 40 column volumes of assay buffer,
5 bound peptide was eluted from the column at approximately 1 M GuHCl with a 4 ml gradient of 0-2 M GuHCl in assay buffer. Fractions were concentrated and desalted. Winston, R. L. and Fitzgerald, M. C. (1998) *Anal Biochem* 262, 83-85. MALDI-MS analysis of individual fractions was used to monitor and characterize peptide elution.

10 *Chemical Synthesis of Boc-Ala-O-Gly.* To prepare the depsipeptide (a peptide that contains an amide to ester substitution), the succinimide ester of Boc-Ala-OH (Boc-Ala-OSu) was reacted with a 2.5 molar excess of glycolic acid in the presence of diisopropylethylamine (DIEA) and methylene chloride, under argon. After 12 hours, the reaction was neutralized with 1 M HCl, and
15 extracted with ethyl acetate. The desired product was isolated by flash chromatography. The purity and identity of the depsipeptide was established by 1 H-NMR and electrospray ionization mass spectrometry (ESI-MS).

Incorporation of Boc-Ala-O-Gly in SPPS. For incorporation into the Dpn polypeptide chain, Boc-Ala-O-Gly (0.25 mmol, 250 μ l of 1 M oil in DMF) was
20 preactivated for 1 hour with DIC (0.25 mmol, 39 μ l) and N-hydroxybenzotriazole (HOBt) (0.25 mmol, 34 mg) in DMF (311 μ l) and used for five consecutive cycles. 125 μ l of the preactivated depsipeptide was then coupled to preneutralized resin for 30 minutes.

Cleavage of Ala-O-Gly Libraries. Hydrazine hydrate was added to
25 eluted protein fractions (final concentration of 1 M hydrazine) and immediately diluted with 1 ml of water. Fractions were desalted and concentrated. Winston, R. L. and Fitzgerald, M. C. (1998) *Anal Biochem* 262, 83-85.

Electrophoretic Mobility Shift Assay. Dpn mutant peptides were assayed using a double stranded specific oligonucleotide (top strand: 5'-
30 CGTACGCCGGCCACGCGACAGGGC-3', where the underlined sequence is the Dpn binding site) (SEQ ID NO:2), in the following assay buffer: 20 mM Hepes, 100 mM KCl, 1 mM EDTA, 5 % glycerol, pH 7.6. Samples were electrophoresed on a 10% nondenaturing polyacrylamide gel, and the data were

analyzed as described. Winston, R. L., Millar, D. P., Gottesfeld, J. M., and Kent, S. B. H. (1999) *Biochemistry* 38, 5138-5146.

- Basic helix-loop-helix (bHLH) transcription factors are characterized by a conserved, parallel four helix bundle that recognizes a specific hexanucleotide DNA sequence in the major groove. See T. Littlewood and G. I. Evan, *Helix-loop-helix transcription factors* (Oxford University Press, New York, 1998); S. J. Anthony-Cahill, et al., *Science* 255, 979-983 (1992); T. D. Halazonetis and A. N. KandiL, *Science* 255, 464-466 (1992); C. R. Vinson and K. C. Garcia, *New Biol.* 4, 396-403 (1992). The least characterized region of these proteins is the loop region, which ranges from 5 to 23 amino acids in length, and varies in amino acid content, especially between proteins of different sub-families. See Littlewood & Evan. The structures of six different bHLH domains show that the loop regions display a large degree of structural variation, while the helical and basic regions are nearly superimposable. See A. R. Ferré-D'Amaré, G. C. Prendergast, E. B. Ziff, S. K. Burley, *Nature* 363, 38-45 (1993); A. R. Ferré-D'Amaré, P. Pognonec, R. G. Roeder, S. K. Burley, *EMBO J.* 13, 180-189 (1994); T. Ellenberger, D. Fass, M. Arnaud, S. C. Harrison, *Genes & Dev.* 8, 970-980 (1994); P. C. M. Ma, M. A. Rould, H. Weintraub, C. O. Pabo, *Cell* 77, 451-459 (1994); A. Párraga, L. Bellsollell, A. R. Ferré-D'Amaré, S. K. Burley, *Structure* 6, 661-672 (1998); T. Shimizu, et al., *EMBO J.* 16, 4689-4697 (1997). It was proposed that a minimum loop of five amino acids is necessary to correctly position helices 1 and 2 in the bHLH fold. See Ferré-D'Amaré et al., *Nature* (1993). However, longer loop regions may play more than a structural role, by contributing to DNA binding affinity and/or specificity through phosphate backbone (3, 4, 6) or base-specific interactions. See Ferré-D'Amaré et al., *Nature* (1993); Ferré-D'Amaré et al., *EMBO J.* (1994); Ma, et al., *Cell* (1994); Ferré-D'Amaré et al., *Nature* (1997). Identification of bHLH loop residues that interact with DNA, and the energetic significance of these contacts has yet to be investigated.
- The predicted loop region of the *Drosophila* bHLH protein Deadpan (Dpn) is 12 to 18 amino acids in length. Littlewood & Evan (1998); E. Bier, H. Vaessin, S. Younger-Shepherd, L. Y. Jan, Y. N. Jan, *Genes & Dev.* 6, 2137-2151 (1992); S. Younger-Shepherd, H. Vaessin, E. Bier, L. Y. Jan, Y. N. Jan, *Cell* 70,

911-922 (1992); W. R. Atchley and W. M. Fitch, *Proc. Natl. Acad. Sci. USA* 94, 5172-5176 (1997); S. R. Dawson, D. L. Turner, H. Weintraub, S. M. Parkhurst, *Mol. Cell. Biol.* 15, 6923-6931 (1995). While the location of helix 2 is defined in all bHLH domains by a strictly conserved lysine residue (Lys 83 in Dpn
5 sequence), the precise end of helix 1 is not obvious for bHLH proteins that lack a semi-conserved proline residue. Littlewood & Evan (1998). Based on a recent, systematic classification of bHLH proteins, the predicted location of the helices and loop region of Dpn are shown in FIG. 1A. Atchley & Fitch, *PNAS* (1998).

In this example, protein-DNA recognition by the *Drosophila* basic helix-
10 loop-helix (bHLH) transcription factor Deadpan was probed using combinatorial solid phase peptide synthesis methods. A series of bHLH peptide libraries which modulate amino acid content and length in the loop region were screened with DNA and peptide affinity columns, and analyzed by matrix-assisted laser desorption ionization mass spectrometry. A functional peptide with reduced
15 loop length was found, and Lys 80 was unambiguously identified as the sole loop residue critical for DNA binding. Unnatural amino acids were substituted at this position to assess contributions of the terminal amino group and the alkyl chain length to DNA binding affinity and specificity. This approach provides a powerful alternative to current recombinant DNA methods to identify and probe
20 the energetics of protein-DNA interactions.

Preparation of Deletion Combinatorial Libraries

In order to define the boundary between helix 1 and the loop, and to determine what role, if any, amino acid side chains in the loop region play in DNA binding, a series of four combinatorial bHLH libraries was generated using
25 manual, stepwise solid-phase peptide synthesis (SPPS) methods (See e.g., M. Schnölzer, P. Alewood, A. Jones, D. Alewood, S. B. H. Kent, *Int. J. Pept. Protein Res.* 40, 180-193 (1992)) were employed to prepare the bHLH portion of Dpn (residues 39-102 in (7)). E. Bier, H. Vaessin, S. Younger-Shepherd, L. Y. Jan, Y. N. Jan, *Genes & Dev.* 6, 2137-2151 (1992); S. Younger-Shepherd, H. Vaessin, E. Bier, L. Y. Jan, Y. N. Jan, *Cell* 70, 911-922 (1992). The length of
30 the loop region was systematically reduced in these combinatorial libraries.

A split resin approach was used to introduce successive, single amino acid deletions (SAD) from both the N- and C- terminal ends of the loop region.

Thus, 4-methylbenzhydryl amine polystyrene resin was functionalized with residues comprising helix 2 (83-102), and then split in half for the generation of N- and C-terminal libraries. N-terminal deletions in the loop region sequence were easily introduced to one half of helix 2 resin, by transferring equimolar
5 portions of resin after each amino acid coupling step to a separate vessel where no amino acid coupling took place. To facilitate subsequent mass spectral analysis, resin containing shorter (N SAD-S) and longer (N SAD-L) loop sequences were transferred to separate vessels.

Introducing deletions from the C-terminal end required a different resin
10 shuffling strategy, because peptide synthesis proceeds in the C to N direction. In this case, equimolar portions of helix 2 resin were added to the main reaction vessel after every amino acid coupling. By repeating this process, a mixture of peptides was generated with systematically deleted loop regions originating from the C-terminal end of the loop. Again, resin containing shorter (C SAD-S) and
15 longer (C SAD-L) loops were kept in separate vessels. To complete the synthesis of the bHLH domain, amino acids from helix 1 and the basic region were assembled, in a parallel fashion, on the four existing resin pools.

This chemical approach obviated recombinant DNA techniques, such as plasmid construction, optimization of protein expression and purification, and
20 characterization of individual mutants. Using this approach, 26 bHLH domain variants were generated in a few days, as depicted in FIG. 1B. Note that the WT-Dpn loop sequence is present in the N SAD-L library as depicted in FIG. 1B, and is marked with an arrow. Each component in each library had a unique mass corresponding to a particular mutant bHLH peptide that can be
25 resolved by matrix-assisted laser desorption ionization mass spectrometry (MALDI-MS), as depicted in FIG. 1C. Observed masses were within experimental uncertainty to calculated masses ($\pm 0.1\%$ Da).

Binding Activity of Deletion Libraries with DNA

In order to determine which mutant peptides retain DNA binding activity,
30 each library was passed over a DNA affinity column containing a known Dpn recognition sequence. This Dpn-specific DNA affinity column was prepared as described using complimentary oligonucleotides containing a Dpn recognition sequence: 5'-CGTACGCCGGCACGCGACAGGTCC-3' (SEQ ID NO:1)(top

strand shown, where the underlined sequence is the Dpn binding site). J. T. Kadonaga and R. Tjian, *Proc. Natl. Acad. Sci. USA* 83, 5889-5893 (1986); R.L. Winston, D. P. Miller, J. M. Gottesfeld, S. B. H. Kent, *Biochemistry* 38, 5138-5146 (1999). The loading capacity of the column was determined to be

5 2 nmole/100 μ l of resin using a WT-Dpn standard. A large excess of protein to DNA (80-fold) was used to ensure competition between peptides. This was because within each peptide library, it is likely that a complex mixture of bHLH heterodimers exist. Heterodimers that form unproductive complexes will be selected against during DNA affinity chromatography. High protein

10 concentrations ensure that all possible heterodimer combinations are represented.

A gradient of increasing ionic strength buffer was used to select for high affinity binding peptides. The buffer that was used in DNA affinity selection experiments was prepared from 20 mM Hepes, 1 mM EDTA, and 5% glycerol, at a pH of 7.6. Initial binding was carried out using a buffer containing 100 mM

15 KCl. Elution steps contained increasing KCl concentrations, as indicated in the text and figure captions. Controls were performed to validate that increasing ionic strength competes away weakly bound peptides and selects for high affinity peptides. Equimolar amounts of three Dpn bHLH peptides (WT-Dpn, Dpn(desPA 75, 76), and Dpn(desDPA 74-77)) with a range of binding

20 affinities (K_d s of 2.6 nM, 4.4 nM, and 44 nM, respectively, for the Dpn site oligonucleotide as determined by EMSA) were pooled and subjected to DNA affinity selection. As a result, Dpn mutant peptides were assayed using a double stranded specific oligonucleotide (top strand: 5'-

CGTACGCCGCGCACGCGACAGGGC-3' (SEQ ID NO:1)', where the

25 underlined sequence is the Dpn binding site)(SEQ ID NO:2), and the data were analyzed using the technique of Winston and coworkers. R.L. Winston, D. P. Miller, J. M. Gottesfeld, S. B. H. Kent, *Biochemistry* 38, 5138-5146 (1999).

Fractions from each step were collected, and subjected to concentration and desalting for MALDI-MS analysis. R. L. Winston and M. C. Fitzgerald,

30 *Anal. Biochem.* 262, 83-85 (1998). FIG. 2A shows MALDI mass spectra of eluted fractions. The elution profile from the functional selection of the N SAD-L library eluted with the indicated KCl concentrations. Ion signals corresponding to WT-Dpn and a mutant missing three amino acids from its N

terminal loop (N-3) are marked with arrows. MALDI-MS analysis of eluted fractions reflected the individual activity of each peptide; i.e., weaker binding peptides eluted at lower ionic strength.

Before selection, all components in the mixture displayed roughly equal ion intensities (top spectrum); however, during the course of DNA-affinity selection, only ion signals corresponding to WT-Dpn and a mutant peptide missing three amino acids from its N-terminal loop (N-3) remained. This result suggests that these three amino acids (residues 68-70) represent the final "helical turn of helix 1, and deletion of all three (but not one or two) amino acids restores the proper helix-loop geometry. Thus, Dpn and members of the Dpn family likely share a similar structure to E47 (5), which contains an extra helical turn at the end of helix 1, as compared to other bHLH proteins such as Max (3).

In the other libraries (C SAD-S, C SAD-L, and N SAD-S), no one mutant could compete effectively against WT-Dpn for DNA binding. As a positive control, an equimolar concentration of WT-Dpn was added to each library (except N SAD-L). To corroborate these findings, peptides eluted from the DNA column were also assayed by electrophoretic mobility shift assay (EMSA), as depicted in FIG. 2B. Samples were equilibrated with a specific DNA probe, subjected to EMSA, and visualized by phosphorimage analysis. In FIG. 2B, lane 1 is DNA alone, and Lanes 2-5 correspond to DNA equilibrated with a 1 μ l aliquot from the 0.6 M KCl fraction for each library as indicated. Each lane contains a similar amount of total protein. Note that only N SAD-L contains WT-Dpn. These results are consistent with the MALDI-MS analyses, showing that only the N SAD-L library contains significantly active material. Indeed, with the exception of the N-3 peptide, these results suggest the possibility that absolute length of the loop region is critical for DNA binding.

To further assess loop length, we generated an internal amino acid deletion (IAD) Dpn library where successive, two amino acid deletions were introduced in the center of the loop, as depicted in FIG. 1D. Monitoring DNA affinity selection of this library by MALDI-MS revealed that only one mutant peptide missing two amino acids (desPA) from the center of the loop had activity comparable to WT-Dpn. This result is depicted in FIG. 2C, in which ion signals corresponding to WT-Dpn and a mutant missing two amino acids (desPA) are

marked with arrows. Thus, loop length per se is not critical for function, however, residues at the loop termini are important for DNA binding.

Because bHLH proteins bind DNA as dimers, we constructed a WT-Dpn peptide affinity column to determine if deletions to the loop region interfered with dimerization. The polypeptide H-Cys-Ahx-[WT-Dpn (39-102)] (where Ahx is amino hexanoic acid) was synthesized and purified using the general procedures described by Winston and coworkers (1999), and reacted with pre-swollen Sulfolink resin (Pierce) under conditions suggested by the manufacturer. Tris-2-carboxyethyl phosphine/HCl (10 mM, pH 8.3) was added to the coupling reaction to prevent peptide disulfide formation. The functional substitution of the column was determined by Bradford assay to be approximately 200 μ M.

We then evaluated the binding activity of the column and found that a linear gradient of 0-2 M guanidine hydrochloride (GuHCl) was sufficient to elute a soluble WT-Dpn standard (20). GuHCl was used because solution studies using circular dichroism spectroscopy as a measure of α -helical content showed that Dpn is completely unfolded in the presence of 2 M GuHCl (unpublished observations). The WT-Dpn standard was prepared using WT-Dpn (1 μ M in the following assay buffer: 100 mM KCl, 1 mM EDTA, 20 mM Hepes, 5% glycerol, pH 7.6) and 83 μ g/ml BSA were incubated with 200 μ l of packed bHLH column resin for 30 minutes with gentle agitation. After washing with 40 column volumes of assay buffer, bound peptide was eluted from the column at approximately 1 M GuHCl with a 4 ml gradient of 0-2 M GuHCl in assay buffer. Fractions were concentrated and desalted. R. L. Winston and M. C. Fitzgerald, *Anal. Biochem.* 262, 83-85 (1998). MALDI-MS analysis of individual fractions was used to monitor and characterize peptide elution.

These conditions were used to assay each library (spiked with WT-Dpn as a positive control). MALDI-MS analysis of these selections shows that there is significant loss of ion signal from N SAD and C SAD libraries compared to WT-Dpn, as depicted in FIG. 2D. According to FIG. 2D, an approximately equimolar concentration of soluble WT-Dpn was added to each library (except N SAD-L and IAD, which contain WT-Dpn). These peptide mixtures were incubated with the bHLH column, and the identity of bound peptides was determined by MALDI-MS analysis of desalted and concentrated elution

fractions. Analysis of the IAD library revealed that the desPA mutant is capable of dimerizing with immobilized WT-Dpn. As a result, deletions originating from the ends of the loop are more deleterious to dimerization than a small deletion in the center of the loop.

- 5 Controls were performed to confirm that binding and elution from the bHLH column reflected the specificity of bHLH dimerization. Increasing concentrations of soluble WT-Dpn added to the libraries resulted in MALDI-MS spectra in which only signals corresponding to WT-Dpn were detected, indicating effective competition of WT-Dpn with the mutant peptides.
- 10 Additionally, libraries incubated with a non-related BSA-linked column showed no selection for WT-Dpn.

Modified Peptide Libraries

Amide to Ester Substitution

- In order to probe amino acid content without modulating the length of the
- 15 loop region, another library was prepared in which a modified peptide containing an amide to ester substitution, Ala-O-Gly, was systematically scanned through eleven positions in the loop (Figures 3A and B). To prepare Ala-O-Gly, the succinimide ester of Boc-Ala-OH (Boc-Ala-OSu) was reacted with a 2.5 molar excess of glycolic acid in the presence of diisopropylethylamine (DIEA) and
- 20 methylene chloride, under argon. After 12 hours, the reaction was neutralized with 1 M HCl, and extracted with ethyl acetate. The desired product was isolated by flash chromatography. The purity and identity of the depsipeptide was established by ¹H-NMR and electrospray ionization mass spectrometry (ESI-MS). For incorporation into the Dpn polypeptide chain, Boc-Ala-O-Gly
- 25 was activated as an N-hydroxybenzotriazole (HOBt) ester and then coupled to pre-neutralized resin.

- This use of Ala-O-Gly serves two purposes: 1) it removes the side chains of two adjacent amino acids and 2) it allows selective cleavage of the peptide at the ester backbone linkage, so no external tagging scheme is required. The
- 30 utility of this approach was demonstrated previously with a similar peptide analog unit containing a thioester backbone. T. W. Muir, P. E. Dawson, M. C. Fitzgerald, S. B. H. Kent, *Chem. Biol.* 3, 817-825 (1996).

Chemical synthesis of the **modified amino acid library (MAL)** was accomplished using a resin shuffling procedure. P.E. Dawson, M. C. Fitzgerald, T. W. Muir, S. B. H. Kent, *J. Am. Chem. Soc.* 119, 7917-7927 (1997). The Ala-O-Gly unit was incorporated once per peptide at a unique position within the
5 loop region. This library was passed over the DNA affinity column and bound peptides were eluted with increasing concentrations of KCl as before. To "decode" components in the eluted fractions, the Ala-O-Gly library was cleaved with 1 M hydrazine and then immediately concentrated and desalted for MALDI-MS analysis. R. L. Winston and M. C. Fitzgerald, *Anal. Biochem.* 262, 83-85
10 (1998). This step broke each bHLH domain into two fragments, yielding N- and C-terminal ladders that reveal the exact location of the Ala-O-Gly linkage. FIG. 3A shows the structure of the Ala-O-Gly linker incorporated into the loop region of Dpn (top) and cleavage of the linker with hydrazine (bottom).

FIG. 3B provides a schematic of the position of the Ala-O-Gly linker
15 (■ ■) in the loop region sequence. Only the sequence corresponding to the modified loop region is shown. Cleavage of the library results in the generation of two peptide fragments (between the ■ ■) for each of the eleven bHLH constructs.

FIG. 3C (top) shows a MALDI mass spectrum of the C-terminal ladder
20 generated after decoding a sample that had not been subjected to DNA-affinity selection. Only ion signals from the C-terminal fragments are shown. Ion signals corresponding to bHLH domains with mutated Lys 80 (indicated with arrows) disappear in the fractions eluting from the DNA affinity column, indicating that these peptides were unable to compete effectively for DNA
25 binding in the presence of the other nine peptides.

After DNA selection, MALDI-MS analysis reveals that two mutant peptides, each missing the side chain of Lys 80, could not compete for DNA binding in the presence of the other nine loop mutants. Because the position of the ester linkage differs in these two peptides, the possibility of backbone amide
30 contributions to DNA binding affinity is eliminated. It is conceivable that other basic residues from the loop contribute to DNA binding activity; however, peptides lacking these amino acid side chains (Lys 72, Lys 73, Arg 77) were not selected against, suggesting that Lys 80 makes a significant and specific DNA

contact. The Ala-O-Gly library was also assayed for dimerization with the bHLH peptide affinity column. MALDI-MS analysis shows that none of the Ala-O-Gly mutations affected dimerization activity (data not shown), indicating that decreased DNA binding activity for Lys 80 mutants is a direct consequence of weakened peptide-DNA interactions, as opposed to diminished bHLH dimerization activity.

Unnatural Amino Acid Substitution

To further investigate the nature of this contact, two peptides each containing an unnatural amino acid substitution at position 80 were individually synthesized and characterized (10). M. Schnölzer, P. Alewood, A. Jones, D. Alewood, S. B. H. Kent, *Int. J. Pept. Protein Res.* 40, 180-193 (1992). The first peptide contained norleucine in place of Lys 80 (Nle 80), which leaves the alkyl side chain of lysine intact but deletes the epsilon amino group, as depicted in FIG. 4A. The second peptide contained ornithine in place of Lys 80 (Orn 80), which maintains the terminal amine, but shortens the alkyl side chain by one methylene.

Crude peptides were purified by reversed-phase HPLC and characterized by analytical reversed-phase HPLC and electrospray ionization mass spectrometry. Winston, R. L., Millar, D. P., Gottesfeld, J. M., and Kent, S. B. H. (1999) *Biochemistry* 38, 5138-5146. The observed masses were within experimental uncertainty to the calculated masses (Nle 80: calc = 7666.1 Da, obs = 7666.4 +/- 0.8 Da; Orn 80: calc = 7667.1 Da, obs = 7667.5 +/- 0.5 Da). Purified peptides were individually assayed for DNA binding by EMSA using a specific DNA probe, and apparent dissociation constants (K_d s) were determined for a 24 bp double stranded oligonucleotide containing a known Dpn binding site (Figure 4B). The observed K_d s were 25 nM and 7 nM for the Nle 80 and Orn 80 Dpn mutants, respectively, compared to 2.6 nM for WT-Dpn (9). Thus, the epsilon amino group of Lys 80 contributes -1.3 kcal/mol to DNA binding affinity, consistent with the energy gained through a phosphate contact. D. R. Lesser, M. R. Kurpiewski, L. Jen-Jacobson, *Science* 250, 776-786 (1990); P. C. Newman, D. M. Williams, R. Cosstick, F. Seela, B. A. Connolly, *Biochemistry* 29, 9902-9910 (1990); C. R. Aiken, L. W. McLaughlin, R. I. Gumpert, *J. Biol. Chem.* 266, 19070-19078 (1991). Adding back the terminal amino group, but

shortening the side chain by one methylene partially restores binding activity (-0.6 kcal/mol). Moreover, a three-fold loss in specificity was observed for Dpn Nle 80 compared to WT-Dpn (Figure 4C), as measured by competition with poly dI-dC (a double stranded DNA mimic). Winston, R. L., Millar, D. P., Gottesfeld, J. M., and Kent, S. B. H. (1999) *Biochemistry* 38, 5138-5146. Therefore, the epsilon amine of Lys 80 makes significant contributions to both DNA affinity and specificity.

Herein a combinatorial strategy was presented that provides information about residues critical for protein-DNA and protein-protein interactions within the Dpn bHLH domain. The boundary between a twelve amino acid loop and the adjacent helix was determined, and despite a wide range of loop lengths found throughout the bHLH protein family, only a small deletion to the center of the loop is tolerated in Dpn. Moreover, we demonstrate that the loop region of Dpn is directly involved in DNA binding, providing significant affinity and specificity to Dpn activity. Using the power of synthetic chemistry, novel functional groups were rationally incorporated into the bHLH domain to closely examine the molecular nuances of Dpn-DNA interactions.

The ability to replace key residues involved in protein-protein or protein-DNA recognition with unnatural amino acids provides a powerful tool with which to dissect and probe energetic contributions to molecular recognition. Because most DNA binding domains are within the accessible range of total chemical synthesis (<100 amino acids), the strategy presented here can be readily be adapted to other structural motifs. Another advantage of this method is that chemical synthesis, selection, and MALDI-MS analysis steps are all amenable to automation. Therefore, rapid characterization of a vast number of synthetic protein domains is feasible. We envision variations of this strategy where novel DNA binding modules could be generated through repeated rounds of synthesis, binding, and selection. Alternatively, a minimal protein domain that interacts with a desired target DNA site could be found by incorporating multiple peptide analogues into a protein scaffold. Our approach could also be extended to study full length proteins by incorporating synthetic peptide libraries into recombinant proteins using the expressed protein ligation strategy.

All publications, patents, and patent documents are incorporated by reference herein, as though individually incorporated by reference. The invention has been described with reference to various specific and preferred embodiments and techniques. However, it should be understood that many variations and modifications may be made while remaining within the spirit and scope of the invention.

What is claimed is:

1. A method for determining the interaction between a polypeptide and a binding agent, comprising
 - 5 a) contacting a library of modified polypeptides with a binding agent known to interact with a lead polypeptide to form a library – binding agent mixture, wherein the lead polypeptide has at least one constant region amino acid sequence and a selected domain amino acid sequence, and each member of the library of the
10 modified polypeptides has the same constant region amino acid sequence as the lead polypeptide and each member of the library of modified polypeptides has a modified domain amino acid sequence that is one or more amino acid unit additions, deletions, substitutions, modifications or combinations thereof of the
15 selected domain amino acid sequence of the lead polypeptide; and
 - b) determining the members of the library of modified polypeptides that have bound to the binding agent.
2. A method according to claim 2 further comprising forming the library of
20 modified polypeptides by substituting a second library of peptide fragments for the selected domain of the lead polypeptide wherein the amino acid sequence of each member of the second library is one or more amino acid unit additions, deletions, substitutions, modifications or combinations thereof of the amino acid sequence of the selected domain.
25
3. A method according to claim 1 wherein the second library of peptide fragments is produced by solid phase peptide synthesis.
4. A method according to claim 1 wherein the lead polypeptide comprises at
30 least two constant regions and the selected domain, the selected domain being positioned between the two constant regions.

5. A method according to claim 1 wherein the constant region is produced by solid phase peptide synthesis or by a recombinant expression technique.
6. A method according to claim 1 wherein the library of modified polypeptides is produced by solid phase peptide synthesis or by a combination of solid phase DNA/RNA synthesis and recombinant expression.
7. A method according to claim 2 wherein each member of the second library is covalently bound to the constant region to form each member of the library of modified polypeptides.
8. A method according to claim 2 wherein all members of the second library are simultaneously bound to constant regions to form the library of modified polypeptides.
9. A method according to claim 1 wherein the binding agent is a substrate, DNA sequence, RNA sequence, antigen, antagonist, carbohydrate, lipid, phospholipid, nucleic acid, agonist, inhibitor, protein binding agent, or receptor activator, small pharmaceutical, small peptide or intracellular chemical messenger.
10. A method according to claim 1 wherein the binding agent is immobilized.
11. A method according to claim 1 wherein the binding agent is bound to a solid support.
12. A method according to claim 1 wherein the binding agent is bound to a microsphere.
13. A method according to claim 1 wherein the determining step includes treating the mixture to form a treated group-binding agent mixture and to remove modified polypeptides that are not bound to the binding agent.

14. A method according to claim 13 wherein the treated group – binding agent mixture is analyzed by mass spectrometry to determine the molecular weights of the modified polypeptides bound to the binding agent.
- 5 15. A method according to claim 14 wherein the mass spectrometry is matrix assisted laser desorption ionization mass spectrometry.
16. A method according to claim 2 wherein each peptide fragment of the second library has the sequence of the selected domain except that one or more
10 amino acid units are deleted from the selected domain sequence to produce each peptide fragment.
17. A method according to claim 2 wherein the peptide fragments of the second library all have the same number of peptide units as the selected domain,
15 and one or more conservative or non-conservative amino acid units are substituted for one or more selected peptide unit of the selected domain to form each peptide fragment.
18. A method according to claim 1 wherein the polypeptide is an enzyme,
20 DNA binding protein, RNA binding protein, antibody, G protein, lipoprotein, chemical messenger binding protein.
19. A library of modified polypeptides wherein each member of the library has a constant region amino acid sequence and a modified domain amino acid
25 sequence, and each member is a modification of a lead polypeptide having the constant region amino acid sequence and a selected domain amino acid sequence, and the modified domain is an amino acid unit addition, deletion, substitution or modification of the amino acid sequence of the selected domain.
- 30 20. A library according to claim 19 wherein the selected domain and modified domain are no more than 100 amino acid units in length.

21. A library according to claim 19 wherein the selected domain and modified domain are no more than 50 amino acid units in length.
22. A library according to claim 19 wherein the modified domain contains at least one cleavable non-amide linkage joining at least two of the units of the domain.
23. A library according to claim 22 wherein the cleavable, non-amide linkage is an ester, thioester, carbonate, allyl or nitro, methoxy phenyl amide linkage.
24. A library according to claim 24 wherein the group of modified domains is a second library.
25. A library according to claim 24 wherein the second library includes a final product having a final amino acid sequence and a group of intermediates having amino acid sequences that are one or more deletions from the final amino acid sequence.
26. A library according to claim 19 wherein the second library includes a final product having a final amino acid sequence and a group of intermediates having amino acid sequences that have one or more conservative or non-conservative amino acid units substituted for one of more selected amino acid units of the final amino acid sequence.
27. A library according to claim 24 wherein the second library is produced by a process of solid phase synthesis wherein amino acid units are sequentially reacted together by chemical synthetic techniques and after addition of each amino acid unit, a portion of the resulting intermediate is isolated, and the remaining portion is used as the starting material for addition of a further amino acid unit until the final product is produced.
28. A library according to claim 24 wherein the second library is produced by a process of solid phase synthesis wherein amino acid units are sequentially

reacted together by chemical synthetic techniques and after addition of each amino acid unit, a first portion of the resulting intermediate is isolated, and the remaining portion is used as the starting material for addition of a further amino acid unit, and the formation of a first portion and remaining portion is repeated
5 after each amino acid unit addition until the final product is produced.

29. A library according to claim 26 wherein after addition of a selected amino acid unit, a first portion of the resulting intermediate is isolated, a conservative or non-conservative amino acid unit is reacted with the first portion
10 to form a second intermediate, the corresponding amino acid unit of the final amino acid sequence is added to the remaining portion to form a third intermediate, and the additional amino acid units of the final amino acid sequence are added to both the second intermediate and the third intermediate to form and the remaining portion is used as the starting material for addition of a
15 further amino acid unit until the final product is produced.

30. A third library of DNA or RNA sequences encoding the library of modified polypeptides of claim 19.

20 31. A third library according to claim 30 wherein a fourth library of fragment DNA or RNA sequences encoding a second library of peptide fragments is produced by solid phase synthesis and the members of the fourth library are ligated to the DNA or RNA sequences encoding the constant region of the modified polypeptides.

25

32. A library of expression vectors containing the DNA sequences of claim 30 in appropriate reading frame configuration to be expressed by a host cell.

30 33. A library of expression vectors according to claim 32 in which the DNA sequences have been combined with a promoter sequence to form an expressible gene.

34. A member of the library of vectors according to claim 32 wherein the polypeptide has been selected according to the assay method of claim 1.
35. A method for identifying a polypeptide that binds to a selected binding agent comprising
- 5 a) contacting a library of modified polypeptides with the selected binding agent to form a group— binding agent mixture, wherein the modified polypeptides have amino acid sequences of from about 6 to 12 amino acid units in length and the amino acid units
- 10 within the sequences are randomly varied; and
- b) determining the individual modified polypeptides that have bound to the binding agent.
36. A method according to claim 35 wherein the selected binding agent is a
- 15 substrate, DNA sequence, RNA sequence, antigen, antagonist, carbohydrate, lipid, phospholipid, nucleic acid, agonist, inhibitor, protein binding agent, or receptor activator, small pharmaceutical, small peptide or intracellular chemical messenger.
- 20 37. A method according to claim 1 wherein each member of the library has an altered amino acid sequence in its selected domain relative to a lead polypeptide.
38. A method according to claim 1 wherein the polypeptide interaction with
- 25 the binding agent is reversible affinity binding.
39. A method according to claim 1 wherein the polypeptide interaction with the binding agent is irreversible affinity binding.
- 30 40. A library of modified domains wherein each member of the library has an amino acid sequence that is one or more amino acid unit additions, deletions, substitutions, modifications or combinations thereof of an amino acid sequence

of a the selected domain of a lead polypeptide, the selected domain being no more than 100 amino acid units in length.

41. A library according to claim 40 wherein the selected domain is no more
5 than 50 amino acid units in length.

42. A library according to claim 40 wherein the selected domain is no more than 20 amino acid units in length.

10 43. A library of modified polypeptides according to claim 19 wherein the lead polypeptide has SEQ ID NO:3 and a second library has SEQ ID NO:4 through SEQ ID NO:31.

44. A library of modified polypeptides according to claim 19 wherein the
15 lead polypeptide has SEQ ID NO:3 and a second library has SEQ ID NO:32 through SEQ ID NO:35.

45. A library of modified polypeptides according to claim 19 wherein the
lead polypeptide has SEQ ID NO:3 and a second library has SEQ ID NO 36
20 through SEQ ID No 46.

1/9

*Fig. 1A*

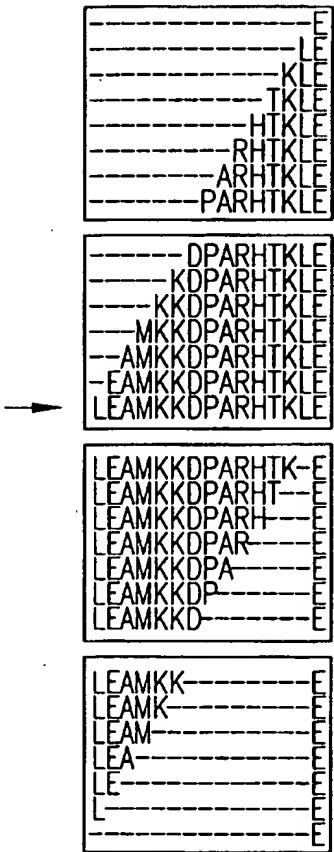


Fig. 1B

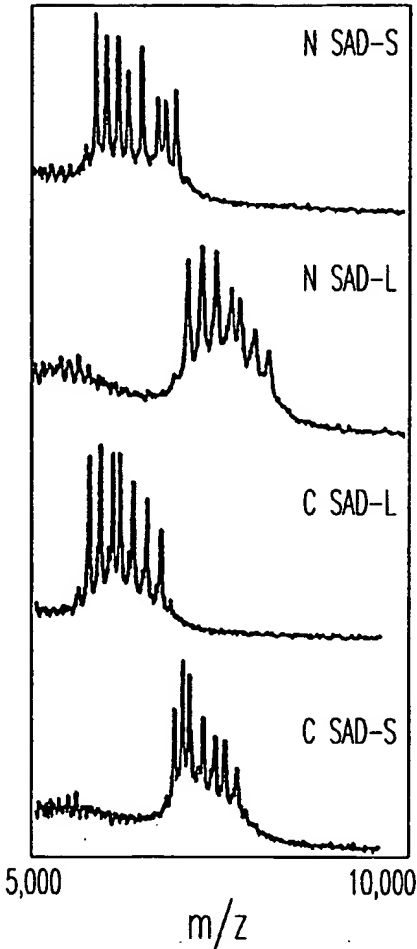
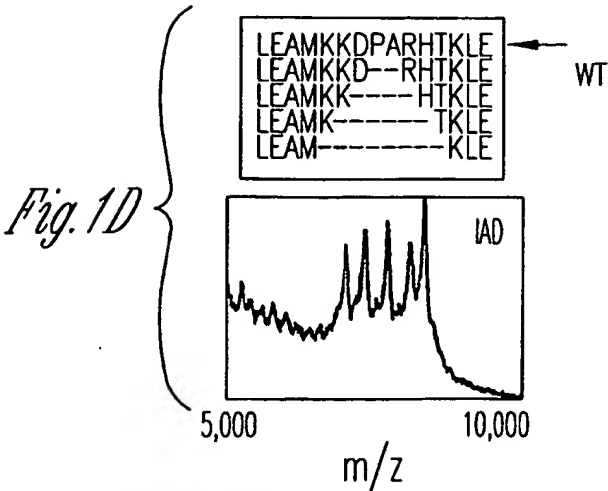
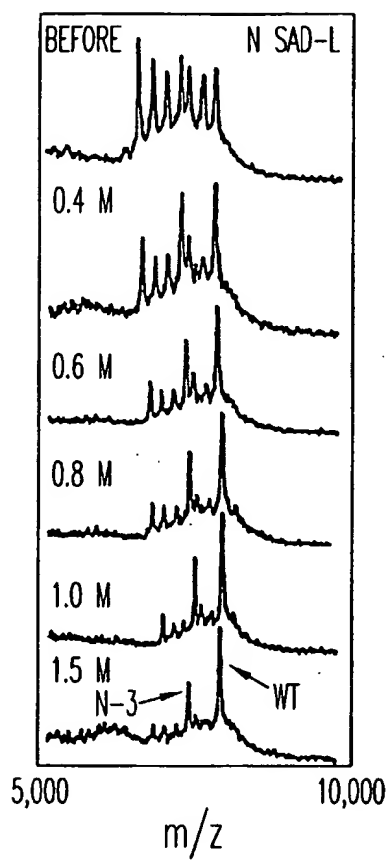
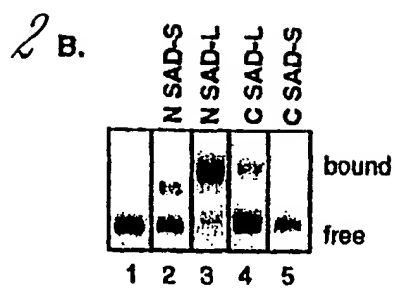


Fig. 1C

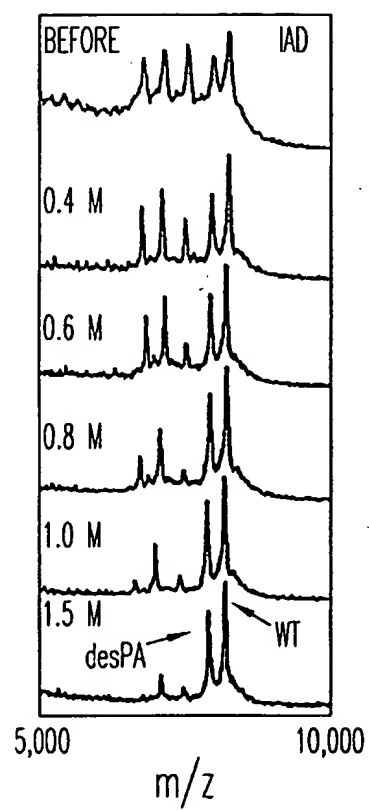
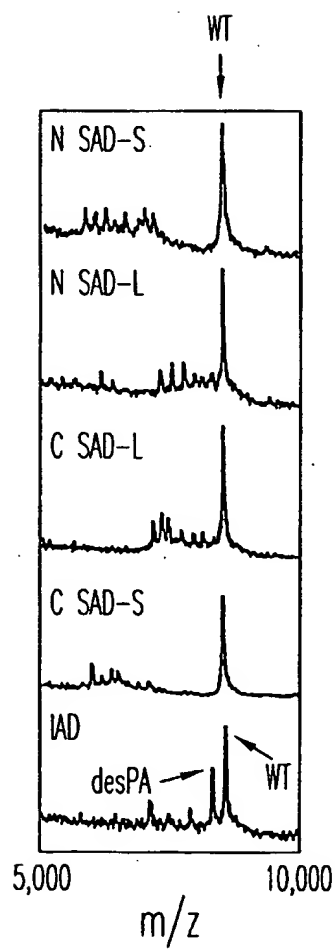


3/9

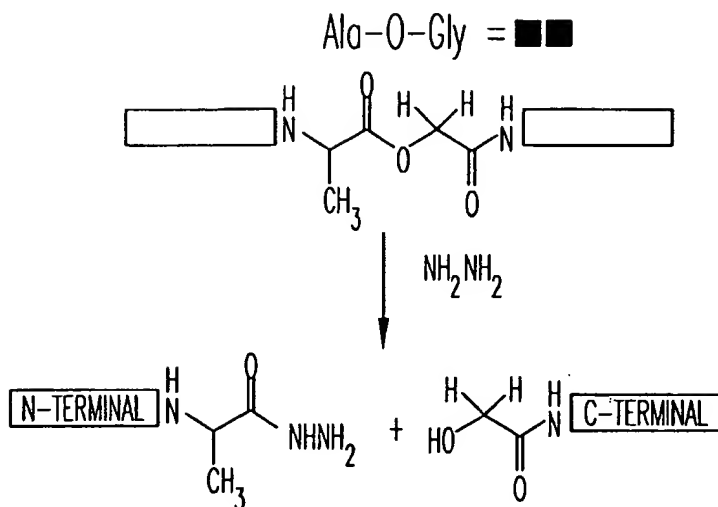
*Fig. 2A*



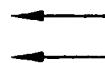
5/9

*Fig. 2C**Fig. 2D*

6/9

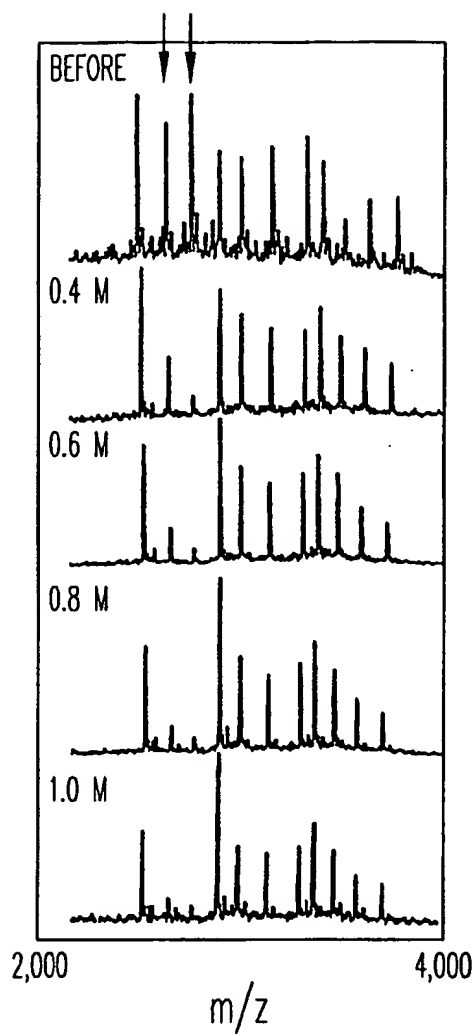
*Fig. 3A*

M	K	K	D	P	A	R	H	T	K	■	■
M	K	K	D	P	A	R	H	T	■	■	E
M	K	K	D	P	A	R	H	■	■	■	L
M	K	K	D	P	A	R	■	■	■	K	L
M	K	K	D	P	A	■	■	■	T	K	L
M	K	K	D	P	■	■	■	H	T	K	L
M	K	K	D	■	■	■	R	H	T	K	L
M	K	K	■	■	■	A	R	H	T	K	L
M	K	■	■	■	P	A	R	H	T	K	L
M	■	■	■	D	P	A	R	H	T	K	L
■	■	■	K	D	P	A	R	H	T	K	L

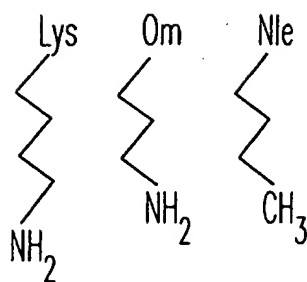
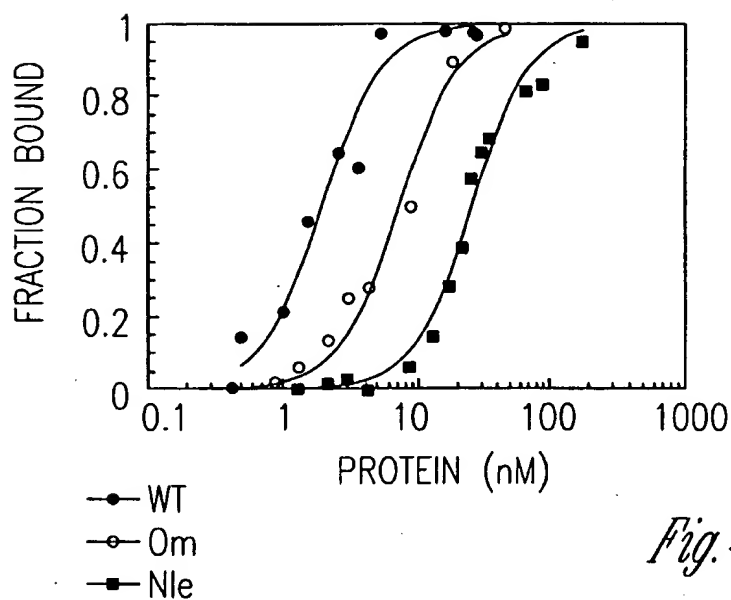
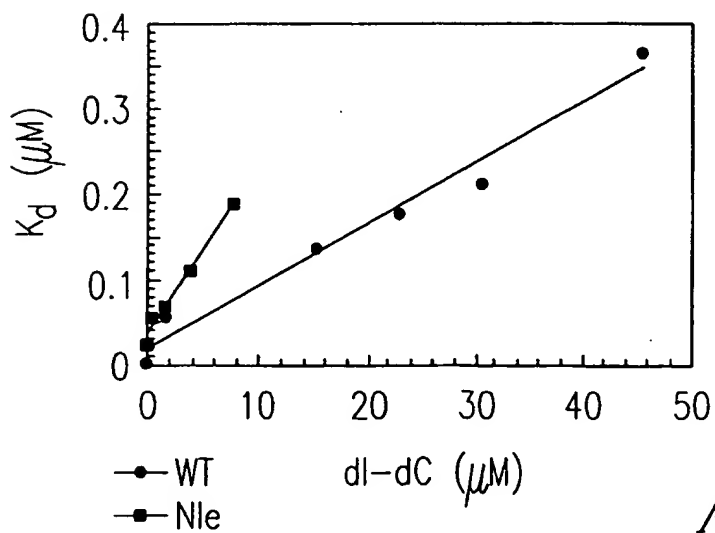
*Fig. 3B*

SUBSTITUTE SHEET (RULE26)

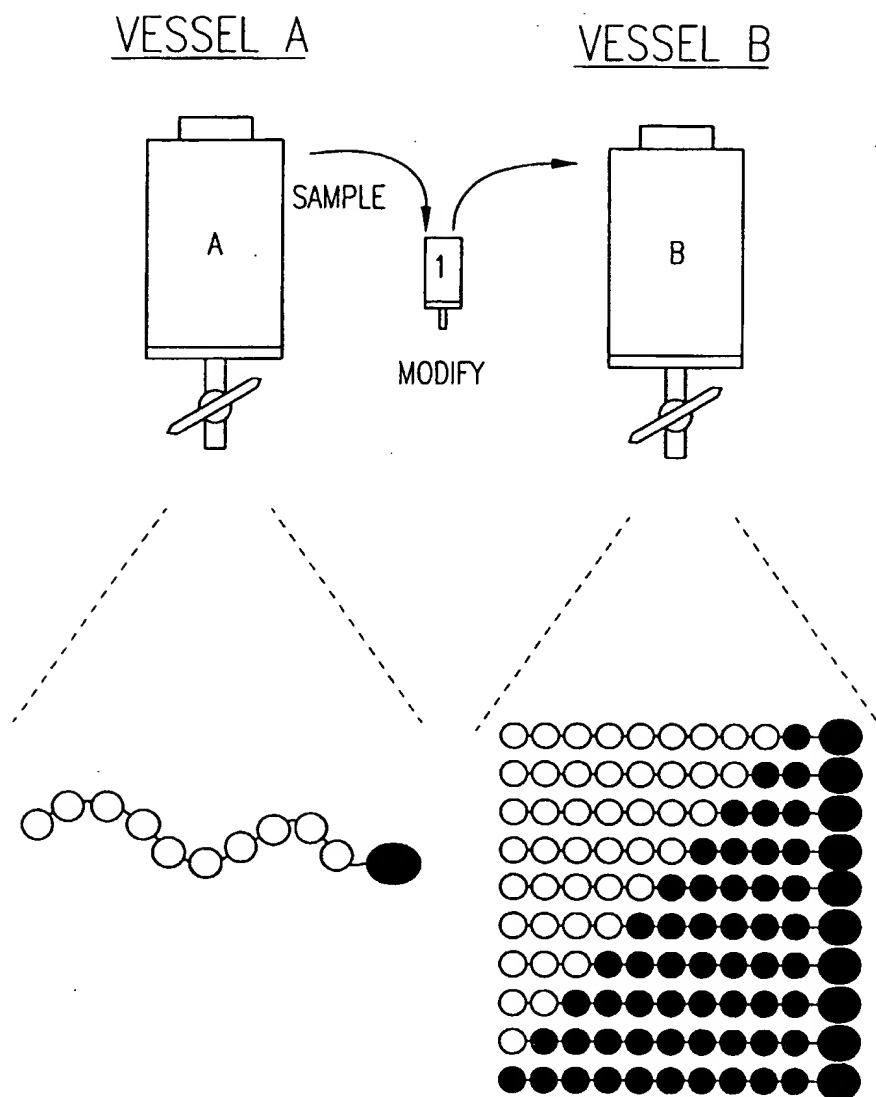
7/9

*Fig. 3C*

8/9

*Fig. 4A**Fig. 4B**Fig. 4C*

9/9

*Fig. 5*

/
SEQUENCE LISTING

<110> The Scripps Research Institute
5 Winston, Rachel L.

<120> Method for determination of
peptide-binding agent interaction

10<130> 1361.002W01

<150> US 60/142,259

<151> 1999-07-02

15<160> 46

<170> FastSEQ for Windows Version 4.0

<210> 1

20<211> 24

<212> DNA

<213> Artificial Sequence

<220>

25<223> An oligonucleotide containing a Dpn recognition
sequence

<400> 1

cgtagcgcgg cagcgacag gtcc

24

30

<210> 2

<211> 23

<212> DNA

<213> Artificial Sequence

35

<220>

<223> An oligonucleotide containing a Dpn recognition
sequence

40<400> 2

cgtagcgcgg cagcgacag ggc

23

2

<210> 3

<211> 64

<212> PRT

<213> Drosophila melanogaster

5

<400> 3

Glu Leu Arg Lys Thr Asn Lys Pro Ile Met Glu Lys Arg Arg Arg Ala
 1 5 10 15
 Arg Ile Asn His Cys Leu Asn Glu Leu Lys Ser Leu Ile Leu Glu Ala
 10 20 25 30
 Met Lys Lys Asp Pro Ala Arg His Thr Lys Leu Glu Lys Ala Asp Ile
 35 40 45
 Leu Glu Met Thr Val Lys His Leu Gln Ser Val Gln Arg Gln Gln Leu
 50 55 60

15

<210> 4

<211> 50

<212> PRT

<213> Artificial Sequence

20

<220>

<223> The bHLH domain of Deadpan with selected deletions

<400> 4

25Glu Leu Arg Lys Thr Asn Lys Pro Ile Met Glu Lys Arg Arg Arg Ala
 1 5 10 15
 Arg Ile Asn His Cys Leu Asn Glu Leu Lys Ser Leu Ile Glu Lys Ala
 20 25 30
 Asp Ile Leu Glu Met Thr Val Lys His Leu Gln Ser Val Gln Arg Gln
 30 35 40 45
 Gln Leu
 50

<210> 5

35<211> 51

<212> PRT

<213> Artificial Sequence

<220>

40<223> The bHLH domain of Deadpan with selected deletions

3

<400> 5

Glu Leu Arg Lys Thr Asn Lys Pro Ile Met Glu Lys Arg Arg Arg Ala

1 5 10 15

Arg Ile Asn His Cys Leu Asn Glu Leu Lys Ser Leu Ile Leu Glu Lys

5 20 25 30

Ala Asp Ile Leu Glu Met Thr Val Lys His Leu Gln Ser Val Gln Arg

35 40 45

Gln Gln Leu

50

10

<210> 6

<211> 52

<212> PRT

<213> Artificial Sequence

15

<220>

<223> The bHLH domain of Deadpan with selected deletions

<400> 6

20Glu Leu Arg Lys Thr Asn Lys Pro Ile Met Glu Lys Arg Arg Arg Ala

1 5 10 15

Arg Ile Asn His Cys Leu Asn Glu Leu Lys Ser Leu Ile Lys Leu Glu

20 25 30

Lys Ala Asp Ile Leu Glu Met Thr Val Lys His Leu Gln Ser Val Gln

25 35 40 45

Arg Gln Gln Leu

50

<210> 7

30<211> 53

<212> PRT

<213> Artificial Sequence

<220>

35<223> The bHLH domain of Deadpan with selected deletions

<400> 7

Glu Leu Arg Lys Thr Asn Lys Pro Ile Met Glu Lys Arg Arg Arg Ala

1 5 10 15

40Arg Ile Asn His Cys Leu Asn Glu Leu Lys Ser Leu Ile Thr Lys Leu

4

20 25 30
 Glu Lys Ala Asp Ile Leu Glu Met Thr Val Lys His Leu Gln Ser Val
 35 40 45
 Gln Arg Gln Gln Leu
 5 50

<210> 8

<211> 54

<212> PRT

10<213> Artificial Sequence

<220>

<223> The bHLH domain of Deadpan with selected deletions

15<400> 8

Glu Leu Arg Lys Thr Asn Lys Pro Ile Met Glu Lys Arg Arg Arg Ala
 1 5 10 15
 Arg Ile Asn His Cys Leu Asn Glu Leu Lys Ser Leu Ile His Thr Lys
 20 25 30
 20Leu Glu Lys Ala Asp Ile Leu Glu Met Thr Val Lys His Leu Gln Ser
 35 40 45
 Val Gln Arg Gln Gln Leu
 50

25<210> 9

<211> 55

<212> PRT

<213> Artificial Sequence

30<220>

<223> The bHLH domain of Deadpan with selected deletions

<400> 9

Glu Leu Arg Lys Thr Asn Lys Pro Ile Met Glu Lys Arg Arg Arg Ala
 35 1 5 10 15
 Arg Ile Asn His Cys Leu Asn Glu Leu Lys Ser Leu Ile Arg His Thr
 20 25 30
 Lys Leu Glu Lys Ala Asp Ile Leu Glu Met Thr Val Lys His Leu Gln
 35 40 45
 40Ser Val Gln Arg Gln Gln Leu

5

50

55

<210> 10

<211> 56

5<212> PRT

<213> Artificial Sequence

<220>

<223> The bHLH domain of Deadpan with selected deletions

10

<400> 10

Glu Leu Arg Lys Thr Asn Lys Pro Ile Met Glu Lys Arg Arg Arg Ala

1

5

10

15

Arg Ile Asn His Cys Leu Asn Glu Leu Lys Ser Leu Ile Ala Arg His

15

20

25

30

Thr Lys Leu Glu Lys Ala Asp Ile Leu Glu Met Thr Val Lys His Leu

35

40

45

Gln Ser Val Gln Arg Gln Gln Leu

50

55

20

<210> 11

<211> 57

<212> PRT

<213> Artificial Sequence

25

<220>

<223> The bHLH domain of Deadpan with selected deletions

<400> 11

30Glu Leu Arg Lys Thr Asn Lys Pro Ile Met Glu Lys Arg Arg Arg Ala

1

5

10

15

Arg Ile Asn His Cys Leu Asn Glu Leu Lys Ser Leu Ile Pro Ala Arg

20

25

30

His Thr Lys Leu Glu Lys Ala Asp Ile Leu Glu Met Thr Val Lys His

35

35

40

45

Leu Gln Ser Val Gln Arg Gln Gln Leu

50

55

<210> 12

40<211> 58

6

<212> PRT

<213> Artificial Sequence

<220>

5<223> The bHLH domain of Deadpan with selected deletions

<400> 12

Glu Leu Arg Lys Thr Asn Lys Pro Ile Met Glu Lys Arg Arg Arg Ala
 1 5 10 15
 10Arg Ile Asn His Cys Leu Asn Glu Leu Lys Ser Leu Ile Asp Pro Ala
 20 25 30
 Arg His Thr Lys Leu Glu Lys Ala Asp Ile Leu Glu Met Thr Val Lys
 35 40 45
 His Leu Gln Ser Val Gln Arg Gln Gln Leu
 15 50 55

<210> 13

<211> 59

<212> PRT

20<213> Artificial Sequence

<220>

<223> The bHLH domain of Deadpan with selected deletions

25<400> 13

Glu Leu Arg Lys Thr Asn Lys Pro Ile Met Glu Lys Arg Arg Arg Ala
 1 5 10 15
 Arg Ile Asn His Cys Leu Asn Glu Leu Lys Ser Leu Ile Lys Asp Pro
 20 25 30
 30Ala Arg His Thr Lys Leu Glu Lys Ala Asp Ile Leu Glu Met Thr Val
 35 40 45
 Lys His Leu Gln Ser Val Gln Arg Gln Gln Leu
 50 55

35<210> 14

<211> 60

<212> PRT

<213> Artificial Sequence

40<220>

<223> The bHLH domain of Deadpan with selected deletions

<400> 14

```

Glu Leu Arg Lys Thr Asn Lys Pro Ile Met Glu Lys Arg Arg Arg Ala
5 1           5           10           15
Arg Ile Asn His Cys Leu Asn Glu Leu Lys Ser Leu Ile Lys Lys Asp
           20           25           30
Pro Ala Arg His Thr Lys Leu Glu Lys Ala Asp Ile Leu Glu Met Thr
           35           40           45
10Val Lys His Leu Gln Ser Val Gln Arg Gln Gln Leu
           50           55           60

```

<210> 15

<211> 61

15<212> PRT

<213> Artificial Sequence

<220>

<223> The bHLH domain of Deadpan with selected deletions

20

<400> 15

```

Glu Leu Arg Lys Thr Asn Lys Pro Ile Met Glu Lys Arg Arg Arg Ala
1           5           10           15
Arg Ile Asn His Cys Leu Asn Glu Leu Lys Ser Leu Ile Met Lys Lys
25           20           25           30
Asp Pro Ala Arg His Thr Lys Leu Glu Lys Ala Asp Ile Leu Glu Met
           35           40           45
Thr Val Lys His Leu Gln Ser Val Gln Arg Gln Gln Leu
           50           55           60

```

30

<210> 16

<211> 62

<212> PRT

<213> Artificial Sequence

35

<220>

<223> The bHLH domain of Deadpan with selected deletions

<400> 16

```

40Glu Leu Arg Lys Thr Asn Lys Pro Ile Met Glu Lys Arg Arg Arg Ala

```

8

1 5 10 15
 Arg Ile Asn His Cys Leu Asn Glu Leu Lys Ser Leu Ile Ala Met Lys
 20 25 30
 Lys Asp Pro Ala Arg His Thr Lys Leu Glu Lys Ala Asp Ile Leu Glu
 5 35 40 45
 Met Thr Val Lys His Leu Gln Ser Val Gln Arg Gln Gln Leu
 50 55 60

<210> 17

10<211> 63

<212> PRT

<213> Artificial Sequence

<220>

15<223> The bHLH domain of Deadpan with selected deletions

<400> 17

Glu Leu Arg Lys Thr Asn Lys Pro Ile Met Glu Lys Arg Arg Arg Ala
 1 5 10 15
 20Arg Ile Asn His Cys Leu Asn Glu Leu Lys Ser Leu Ile Glu Ala Met
 20 25 30
 Lys Lys Asp Pro Ala Arg His Thr Lys Leu Glu Lys Ala Asp Ile Leu
 35 40 45
 Glu Met Thr Val Lys His Leu Gln Ser Val Gln Arg Gln Gln Leu
 25 50 55 60

<210> 18

<211> 63

<212> PRT

30<213> Artificial Sequence

<220>

<223> The bHLH domain of Deadpan with selected deletions

35<400> 18

Glu Leu Arg Lys Thr Asn Lys Pro Ile Met Glu Lys Arg Arg Arg Ala
 1 5 10 15
 Arg Ile Asn His Cys Leu Asn Glu Leu Lys Ser Leu Ile Leu Glu Ala
 20 25 30
 40Met Lys Lys Asp Pro Ala Arg His Thr Lys Glu Lys Ala Asp Ile Leu

35 40 45
 Glu Met Thr Val Lys His Leu Gln Ser Val Gln Arg Gln Gln Leu
 50 55 60

5<210> 19

<211> 62

<212> PRT

<213> Artificial Sequence

10<220>

<223> The bHLH domain of Deadpan with selected deletions

<400> 19

Glu Leu Arg Lys Thr Asn Lys Pro Ile Met Glu Lys Arg Arg Arg Ala
 15 1 5 10 15
 Arg Ile Asn His Cys Leu Asn Glu Leu Lys Ser Leu Ile Leu Glu Ala
 20 25 30
 Met Lys Lys Asp Pro Ala Arg His Thr Glu Lys Ala Asp Ile Leu Glu
 35 40 45
 20Met Thr Val Lys His Leu Gln Ser Val Gln Arg Gln Gln Leu
 50 55 60

<210> 20

<211> 61

25<212> PRT

<213> Artificial Sequence

<220>

<223> The bHLH domain of Deadpan with selected deletions

30

<400> 20

Glu Leu Arg Lys Thr Asn Lys Pro Ile Met Glu Lys Arg Arg Arg Ala
 1 5 10 15
 Arg Ile Asn His Cys Leu Asn Glu Leu Lys Ser Leu Ile Leu Glu Ala
 35 20 25 30
 Met Lys Lys Asp Pro Ala Arg His Glu Lys Ala Asp Ile Leu Glu Met
 35 40 45
 Thr Val Lys His Leu Gln Ser Val Gln Arg Gln Gln Leu
 50 55 60

40

10

<210> 21

<211> 60

<212> PRT

<213> Artificial Sequence

5

<220>

<223> The bHLH domain of Deadpan with selected deletions

<400> 21

10Glu Leu Arg Lys Thr Asn Lys Pro Ile Met Glu Lys Arg Arg Arg Ala

1 5 10 15

Arg Ile Asn His Cys Leu Asn Glu Leu Lys Ser Leu Ile Leu Glu Ala

20 25 30

Met Lys Lys Asp Pro Ala Arg Glu Lys Ala Asp Ile Leu Glu Met Thr

15 35 40 45

Val Lys His Leu Gln Ser Val Gln Arg Gln Gln Leu

50 55 60

<210> 22

20<211> 59

<212> PRT

<213> Artificial Sequence

<220>

25<223> The bHLH domain of Deadpan with selected deletions

<400> 22

Glu Leu Arg Lys Thr Asn Lys Pro Ile Met Glu Lys Arg Arg Arg Ala

1 5 10 15

30Arg Ile Asn His Cys Leu Asn Glu Leu Lys Ser Leu Ile Leu Glu Ala

20 25 30

Met Lys Lys Asp Pro Ala Glu Lys Ala Asp Ile Leu Glu Met Thr Val

35 40 45

Lys His Leu Gln Ser Val Gln Arg Gln Gln Leu

35 50 55

<210> 23

<211> 58

<212> PRT

40<213> Artificial Sequence

<220>

<223> The bHLH domain of Deadpan with selected deletions

<400> 23

5Glu Leu Arg Lys Thr Asn Lys Pro Ile Met Glu Lys Arg Arg Arg Ala
1 5 10 15
Arg Ile Asn His Cys Leu Asn Glu Leu Lys Ser Leu Ile Leu Glu Ala
20 25 30
Met Lys Lys Asp Pro Glu Lys Ala Asp Ile Leu Glu Met Thr Val Lys
10 35 40 45
His Leu Gln Ser Val Gln Arg Gln Gln Leu
50 55

<210> 24

15<211> 57

<212> PRT

<213> Artificial Sequence

<220>

20<223> The bHLH domain of Deadpan with selected deletions

<400> 24

Glu Leu Arg Lys Thr Asn Lys Pro Ile Met Glu Lys Arg Arg Arg Ala
 1 5 10 15
 25Arg Ile Asn His Cys Leu Asn Glu Leu Lys Ser Leu Ile Leu Glu Ala
 20 25 30
 Met Lys Lys Asp Glu Lys Ala Asp Ile Leu Glu Met Thr Val Lys His
 35 40 45
 Leu Gln Ser Val Gln Arg Gln Gln Leu
 50 55

<210> 25

<211> 56

<212> PRT

35<213> Artificial Sequence

<220>

<223> The bHLH domain of Deadpan with selected deletions

40<400> 25

12

Glu Leu Arg Lys Thr Asn Lys Pro Ile Met Glu Lys Arg Arg Arg Ala
 1 5 10 15
 Arg Ile Asn His Cys Leu Asn Glu Leu Lys Ser Leu Ile Leu Glu Ala
 20 25 30
 5Met Lys Lys Glu Lys Ala Asp Ile Leu Glu Met Thr Val Lys His Leu
 35 40 45
 Gln Ser Val Gln Arg Gln Gln Leu
 50 55

10<210> 26

<211> 55

<212> PRT

<213> Artificial Sequence

15<220>

<223> The bHLH domain of Deadpan with selected deletions

<400> 26

Glu Leu Arg Lys Thr Asn Lys Pro Ile Met Glu Lys Arg Arg Arg Ala
 20 1 5 10 15
 Arg Ile Asn His Cys Leu Asn Glu Leu Lys Ser Leu Ile Leu Glu Ala
 20 25 30
 Met Lys Glu Lys Ala Asp Ile Leu Glu Met Thr Val Lys His Leu Gln
 35 40 45
 25 Ser Val Gln Arg Gln Gln Leu
 50 55

<210> 27

<211> 54

30<212> PRT

<213> Artificial Sequence

<220>

<223> The bHLH domain of Deadpan with selected deletions

35

<400> 27

Glu Leu Arg Lys Thr Asn Lys Pro Ile Met Glu Lys Arg Arg Arg Ala
 1 5 10 15
 Arg Ile Asn His Cys Leu Asn Glu Leu Lys Ser Leu Ile Leu Glu Ala
 40 20 25 30

13

Met Glu Lys Ala Asp Ile Leu Glu Met Thr Val Lys His Leu Gln Ser

35

40

45

Val Gln Arg Gln Gln Leu

50

5

<210> 28

<211> 53

<212> PRT

<213> Artificial Sequence

10

<220>

<223> The bHLH domain of Deadpan with selected deletions

<400> 28

15Glu Leu Arg Lys Thr Asn Lys Pro Ile Met Glu Lys Arg Arg Arg Ala

1

5

10

15

Arg Ile Asn His Cys Leu Asn Glu Leu Lys Ser Leu Ile Leu Glu Ala

20

25

30

Glu Lys Ala Asp Ile Leu Glu Met Thr Val Lys His Leu Gln Ser Val

20

35

40

45

Gln Arg Gln Gln Leu

50

<210> 29

25<211> 52

<212> PRT

<213> Artificial Sequence

<220>

30<223> The bHLH domain of Deadpan with selected deletions

<400> 29

Glu Leu Arg Lys Thr Asn Lys Pro Ile Met Glu Lys Arg Arg Arg Ala

1

5

10

15

35Arg Ile Asn His Cys Leu Asn Glu Leu Lys Ser Leu Ile Leu Glu Glu

20

25

30

Lys Ala Asp Ile Leu Glu Met Thr Val Lys His Leu Gln Ser Val Gln .

35

40

45

Arg Gln Gln Leu

40 50

14

<210> 30

<211> 51

<212> PRT

<213> Artificial Sequence

5

<220>

<223> The bHLH domain of Deadpan with selected deletions

<400> 30

10Glu Leu Arg Lys Thr Asn Lys Pro Ile Met Glu Lys Arg Arg Arg Ala

1

5

10

15

Arg Ile Asn His Cys Leu Asn Glu Leu Lys Ser Leu Ile Leu Glu Lys

20

25

30

Ala Asp Ile Leu Glu Met Thr Val Lys His Leu Gln Ser Val Gln Arg

15

35

40

45

Gln Gln Leu

50

<210> 31

20<211> 50

<212> PRT

<213> Artificial Sequence

<220>

25<223> The bHLH domain of Deadpan with selected deletions

<400> 31

Glu Leu Arg Lys Thr Asn Lys Pro Ile Met Glu Lys Arg Arg Arg Ala

1

5

10

15

30Arg Ile Asn His Cys Leu Asn Glu Leu Lys Ser Leu Ile Glu Lys Ala

20

25

30

Asp Ile Leu Glu Met Thr Val Lys His Leu Gln Ser Val Gln Arg Gln

35

40

45

Gln Leu

35 50

<210> 32

<211> 62

<212> PRT

40<213> Artificial Sequence

15

<220>

<223> The bHLH domain of Deadpan with selected deletions

<400> 32

5Glu Leu Arg Lys Thr Asn Lys Pro Ile Met Glu Lys Arg Arg Arg Ala
 1 5 10 15
 Arg Ile Asn His Cys Leu Asn Glu Leu Lys Ser Leu Ile Leu Glu Ala
 20 25 30
 Met Lys Lys Asp Arg His Thr Lys Leu Glu Lys Ala Asp Ile Leu Glu
 10 35 40 45
 Met Thr Val Lys His Leu Gln Ser Val Gln Arg Gln Gln Leu
 50 55 60

<210> 33

15<211> 60

<212> PRT

<213> Artificial Sequence

<220>

20<223> The bHLH domain of Deadpan with selected deletions

<400> 33

Glu Leu Arg Lys Thr Asn Lys Pro Ile Met Glu Lys Arg Arg Arg Ala
 1 5 10 15
 25Arg Ile Asn His Cys Leu Asn Glu Leu Lys Ser Leu Ile Leu Glu Ala
 20 25 30
 Met Lys Lys His Thr Lys Leu Glu Lys Ala Asp Ile Leu Glu Met Thr
 35 40 45
 Val Lys His Leu Gln Ser Val Gln Arg Gln Gln Leu
 30 50 55 60

<210> 34

<211> 58

<212> PRT

35<213> Artificial Sequence

<220>

<223> The bHLH domain of Deadpan with selected deletions

40<400> 34

16

Glu Leu Arg Lys Thr Asn Lys Pro Ile Met Glu Lys Arg Arg Arg Ala
 1 5 10 15
 Arg Ile Asn His Cys Leu Asn Glu Leu Lys Ser Leu Ile Leu Glu Ala
 20 25 30
 5Met Lys Thr Lys Leu Glu Lys Ala Asp Ile Leu Glu Met Thr Val Lys
 35 40 45
 His Leu Gln Ser Val Gln Arg Gln Gln Leu
 50 55

10<210> 35

<211> 56

<212> PRT

<213> Artificial Sequence

15<220>

<223> The bHLH domain of Deadpan with selected deletions

<400> 35

Glu Leu Arg Lys Thr Asn Lys Pro Ile Met Glu Lys Arg Arg Arg Ala
 20 1 5 10 15
 Arg Ile Asn His Cys Leu Asn Glu Leu Lys Ser Leu Ile Leu Glu Ala
 20 25 30
 Met Lys Leu Glu Lys Ala Asp Ile Leu Glu Met Thr Val Lys His Leu
 35 40 45
 25Gln Ser Val Gln Arg Gln Gln Leu
 50 55

<210> 36

<211> 64

30<212> PRT

<213> Artificial Sequence

<220>

<223> The bHLH domain of Deadpan with the incorporation
 35 of an Ala-O-Gly linker

<221> SITE

<222> (43)...(44)

<223> An ester bond

40

17

<400> 36
 Glu Leu Arg Lys Thr Asn Lys Pro Ile Met Glu Lys Arg Arg Arg Ala
 1 5 10 15
 Arg Ile Asn His Cys Leu Asn Glu Leu Lys Ser Leu Ile Leu Glu Ala
 5 20 25 30
 Met Lys Lys Asp Pro Ala Arg His Thr Lys Ala Gly Lys Ala Asp Ile
 35 40 45
 Leu Glu Met Thr Val Lys His Leu Gln Ser Val Gln Arg Gln Gln Leu
 50 55 60

10

<210> 37
 <211> 64
 <212> PRT
 <213> Artificial Sequence

15

<220>
 <223> The bHLH domain of Deadpan with the incorporation
 of an Ala-O-Gly linker

20<221> SITE

<222> (42)...(43)
 <223> An ester bond

<400> 37
 25Glu Leu Arg Lys Thr Asn Lys Pro Ile Met Glu Lys Arg Arg Arg Ala
 1 5 10 15
 Arg Ile Asn His Cys Leu Asn Glu Leu Lys Ser Leu Ile Leu Glu Ala
 20 25 30
 Met Lys Lys Asp Pro Ala Arg His Thr Ala Gly Glu Lys Ala Asp Ile
 30 35 40 45
 Leu Glu Met Thr Val Lys His Leu Gln Ser Val Gln Arg Gln Gln Leu
 50 55 60

<210> 38
 35<211> 64
 <212> PRT
 <213> Artificial Sequence

<220>
 40<223> The bHLH domain of Deadpan with the incorporation

18

of an Ala-O-Gly linker

<221> SITE

<222> (41)...(42)

5<223> An ester bond

<400> 38

Glu	Leu	Arg	Lys	Thr	Asn	Lys	Pro	Ile	Met	Glu	Lys	Arg	Arg	Arg	Ala
1				5					10					15	
10	Arg	Ile	Asn	His	Cys	Leu	Asn	Glu	Leu	Lys	Ser	Leu	Ile	Leu	Glu
			20						25					30	
	Met	Lys	Lys	Asp	Pro	Ala	Arg	His	Ala	Gly	Leu	Glu	Lys	Ala	Asp
			35						40					45	
	Leu	Glu	Met	Thr	Val	Lys	His	Leu	Gln	Ser	Val	Gln	Arg	Gln	Leu
15		50						55					60		

<210> 39

<211> 64

<212> PRT

20<213> Artificial Sequence

<220>

<223> The bHLH domain of Deadpan with the incorporation
of an Ala-O-Gly linker

25

<221> SITE

<222> (40)...(41)

<223> An ester bond

30<400> 39

Glu	Leu	Arg	Lys	Thr	Asn	Lys	Pro	Ile	Met	Glu	Lys	Arg	Arg	Arg	Ala
1				5					10					15	
	Arg	Ile	Asn	His	Cys	Leu	Asn	Glu	Leu	Lys	Ser	Leu	Ile	Leu	Glu
			20						25					30	
35	Met	Lys	Lys	Asp	Pro	Ala	Arg	Ala	Gly	Lys	Leu	Glu	Lys	Ala	Asp
			35						40					45	
	Leu	Glu	Met	Thr	Val	Lys	His	Leu	Gln	Ser	Val	Gln	Arg	Gln	Leu
		50						55					60		

40<210> 40

19

<211> 64

<212> PRT

<213> Artificial Sequence

5<220>

<223> The bHLH domain of Deadpan with the incorporation
of an Ala-O-Gly linker

<221> SITE

10<222> (39)...(40)

<223> An ester bond

<400> 40

	Glu	Leu	Arg	Lys	Thr	Asn	Lys	Pro	Ile	Met	Glu	Lys	Arg	Arg	Arg	Ala
15	1				5				10					15		
	Arg	Ile	Asn	His	Cys	Leu	Asn	Glu	Leu	Lys	Ser	Leu	Ile	Leu	Glu	Ala
			20					25					30			
	Met	Lys	Lys	Asp	Pro	Ala	Ala	Gly	Thr	Lys	Leu	Glu	Lys	Ala	Asp	Ile
			35					40					45			
20	Leu	Glu	Met	Thr	Val	Lys	His	Leu	Gln	Ser	Val	Gln	Arg	Gln	Gln	Leu
			50					55					60			

<210> 41

<211> 64

25<212> PRT

<213> Artificial Sequence

<220>

<223> The bHLH domain of Deadpan with the incorporation
30 of an Ala-O-Gly linker

<221> SITE

<222> (38)...(39)

<223> An ester bond

35

<400> 41

	Glu	Leu	Arg	Lys	Thr	Asn	Lys	Pro	Ile	Met	Glu	Lys	Arg	Arg	Arg	Ala
	1				5				10					15		
	Arg	Ile	Asn	His	Cys	Leu	Asn	Glu	Leu	Lys	Ser	Leu	Ile	Leu	Glu	Ala
40				20					25					30		

20

Met Lys Lys Asp Pro Ala Gly His Thr Lys Leu Glu Lys Ala Asp Ile

35

40

45

Leu Glu Met Thr Val Lys His Leu Gln Ser Val Gln Arg Gln Gln Leu

50

55

60

5

<210> 42

<211> 64

<212> PRT

<213> Artificial Sequence

10

<220>

<223> The bHLH domain of Deadpan with the incorporation
of an Ala-O-Gly linker

15<221> SITE

<222> (37)...(38)

<223> An ester bond

<400> 42

20Glu Leu Arg Lys Thr Asn Lys Pro Ile Met Glu Lys Arg Arg Arg Ala

1

5

10

15

Arg Ile Asn His Cys Leu Asn Glu Leu Lys Ser Leu Ile Leu Glu Ala

20

25

30

Met Lys Lys Asp Ala Gly Arg His Thr Lys Leu Glu Lys Ala Asp Ile

25

35

40

45

Leu Glu Met Thr Val Lys His Leu Gln Ser Val Gln Arg Gln Gln Leu

50

55

60

<210> 43

30<211> 64

<212> PRT

<213> Artificial Sequence

<220>

35<223> The bHLH domain of Deadpan with the incorporation
of an Ala-O-Gly linker

<221> SITE

<222> (36)...(37)

40<223> An ester bond

21

<400> 43

Glu Leu Arg Lys Thr Asn Lys Pro Ile Met Glu Lys Arg Arg Arg Ala

1 5 10 15

Arg Ile Asn His Cys Leu Asn Glu Leu Lys Ser Leu Ile Leu Glu Ala

5 20 25 30

Met Lys Lys Ala Gly Ala Arg His Thr Lys Leu Glu Lys Ala Asp Ile

35 40 45

Leu Glu Met Thr Val Lys His Leu Gln Ser Val Gln Arg Gln Gln Leu

50 55 60

10

<210> 44

<211> 64

<212> PRT

<213> Artificial Sequence

15

<220>

<223> The bHLH domain of Deadpan with the incorporation
of an Ala-O-Gly linker

20<221> SITE

<222> (35)...(36)

<223> An ester bond

<400> 44

25Glu Leu Arg Lys Thr Asn Lys Pro Ile Met Glu Lys Arg Arg Arg Ala

1 5 10 15

Arg Ile Asn His Cys Leu Asn Glu Leu Lys Ser Leu Ile Leu Glu Ala

20 25 30

Met Lys Ala Gly Pro Ala Arg His Thr Lys Leu Glu Lys Ala Asp Ile

30 35 40 45

Leu Glu Met Thr Val Lys His Leu Gln Ser Val Gln Arg Gln Gln Leu

50 55 60

<210> 45

35<211> 64

<212> PRT

<213> Artificial Sequence

<220>

40<223> The bHLH domain of Deadpan with the incorporation

22

of an Ala-O-Gly linker

<221> SITE

<222> (34)...(35)

5<223> An ester bond

<400> 45

Glu	Leu	Arg	Lys	Thr	Asn	Lys	Pro	Ile	Met	Glu	Lys	Arg	Arg	Arg	Ala
1				5				10					15		
10	Arg	Ile	Asn	His	Cys	Leu	Asn	Glu	Leu	Lys	Ser	Leu	Ile	Leu	Glu
			20					25					30		
	Met	Ala	Gly	Asp	Pro	Ala	Arg	His	Thr	Lys	Leu	Glu	Lys	Ala	Asp
		35					40					45			Ile
	Leu	Glu	Met	Thr	Val	Lys	His	Leu	Gln	Ser	Val	Gln	Arg	Gln	Gln
15		50					55					60			Leu

<210> 46

<211> 64

<212> PRT

20<213> Artificial Sequence

<220>

<223> The bHLH domain of Deadpan with the incorporation
of an Ala-O-Gly linker

25

<221> SITE

<222> (33)...(34)


<223> An ester bond

30<400> 46

Glu	Leu	Arg	Lys	Thr	Asn	Lys	Pro	Ile	Met	Glu	Lys	Arg	Arg	Arg	Ala
1				5				10					15		
	Arg	Ile	Asn	His	Cys	Leu	Asn	Glu	Leu	Lys	Ser	Leu	Ile	Leu	Glu
			20					25					30		
35	Ala	Gly	Lys	Asp	Pro	Ala	Arg	His	Thr	Lys	Leu	Glu	Lys	Ala	Asp
		35					40					45			Ile
	Leu	Glu	Met	Thr	Val	Lys	His	Leu	Gln	Ser	Val	Gln	Arg	Gln	Gln
		50					55					60			Leu

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/18335

A. CLASSIFICATION OF SUBJECT MATTER IPC(7) : G01N 33/53; C07K 14/00; C12N 15/11 US CL : 435/7.1; 530/324, 325; 536/23.1 According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) U.S. : 435/7.1; 530/324, 325; 536/23.1 Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) cas, west		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	HOMPSON, L.A. et al. Synthesis and applications of small molecule libraries. Chem. Rev. 1996, Vol. 96, No. 1, pages 555-600, especially pages 560 and 562-563.	1-45
Y	US 5,789,538 A (REBAR et al) 04 August 1998, col. 2, line 35 up to col. 12, line 10.	1-45
Y	WU, H. et al. Building zinc fingers by selection: toward a therapeutic application. Proc. Natl. Acad. Sci. January 1999, Vol. 92, pages 344-348, entire document.	1-45
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> See patent family annex.		
* Special categories of cited documents: *A* document defining the general state of the art which is not considered to be of particular relevance *B* earlier document published on or after the international filing date *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) *O* document referring to an oral disclosure, use, exhibition or other means *P* document published prior to the international filing date but later than the priority date claimed	*T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art *A* document member of the same patent family	
Date of the actual completion of the international search 27 SEPTEMBER 2000		Date of mailing of the international search report 27 OCT 2000
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230		Authorized officer:  T. WESSENDORF Telephone No. (703) 308-0196

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JS00/18335

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,P	WINSTON, R.L. et al. Rapid identification of key amino -acid-DNA contacts through combinatorial peptide synthesis. Chemistry & Biology. 14 March 2000, Vol. 7, pages 245-251, see entire document.	1-45
Y	WINSTON, R.L. et al. Characterization of the DNA binding properties of the bHLH domain of deadpan to single and tandem sites. Biochemistry. 03 March 1999, Vol. 38, No. 16, pages 5138-5146, see entire document.	1-45
Y,P	US 6,007,988 A (CHOO et al) 28 December 1999, see entire document.	1-45

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/18335

Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Please See Extra Sheet.

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☒ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US00/18335

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING

This ISA found multiple inventions as follows:

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

Group I, claim(s) 1-29 and 37-39 and 43-45, drawn to a method for determining the interaction between a polypeptide and a binding agent and library of modified peptides.

Group II, claim(s) 30-34, drawn to third library of DNA or RNA sequences.

Group III, claim(s) 35-36, drawn to a method of identifying a peptide that binds to an agent.

Group IV, claims 40-42, drawn to a modified polypeptide domains.

The invention of Groups I, II and IV are not so linked as to form a single inventive concept under PCT Rule 13.1 because Group I library of peptide containing the basic units of amino acids lack the same or corresponding structure of the nucleic acids comprised in the library of DNA or RNA so as to form a single inventive concept. Also, because of the degenerate codes of DNA or RNA different library of peptides could result. Group IV contains structure wherein the amino acids of the parent library peptide of Group I has been modified individually or in combinations therefore, would contain a structure and modes of actions or effects different from that of the parent library of peptides of Group I. The inventions of Groups I, II and IV do not form a single inventive concept with a method recited in Group III method, as the library of each of the compounds may be employed for other uses.

This application contains claims directed to more than one species of the generic invention. These species are deemed to lack Unity of Invention because they are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for more than one species to be searched, the appropriate additional search fees must be paid. The species are as follows:

Species of Sequence IDs. 4-46.

Each of these species lack the corresponding structure to form a single inventive concept.

The claims are deemed to correspond to the species listed above in the following manner:

Claims 43-45.

The following claims are generic: e.g., claim 19.